DOCUMENT RESUME

ED 190 104 IR 008 588

AUTHOR Marcus, Richard S.; Reintjes, J. Francis
TITLE Experiments and Analysis on a Computer Interface to an Information-Retrieval Network.
INSTITUTION Massachusetts Inst. of Tech., Cambridge. Laboratory for Information and Decision Systems.
SPONS AGENCY National Science Foundation. Washington, D.C. Div. of Information Science and Technology.
REPORT NO LIDS-R-900
PUB DATE Apr 79
GRANT NSF-IST-76-82117
NOTE 133p.

EDRS PRICE MF01/PC06 Plus Postage.
DESCRIPTORS Computers: Data Bases: *Information Networks: *Information Retrieval: Information Systems: *Man Machine Systems: *Online Systems

ABSTRACT

A primary goal of this project was to develop an interface that would provide direct access for inexperienced users to existing online bibliographic information retrieval networks. The experiment tested the concept of a virtual-system mode of access to a network of heterogeneous interactive retrieval systems and databases. An experimental translating computer interface named CONIT, that enables the virtual-system mode, was developed as a research test vehicle. The interface was designed to make the basic functions of three different bibliographic retrieval systems easy to use, even by inexperienced end users, by providing a simplified common command language coupled with extensive online instruction. Analysis of controlled experiments with end users indicates the probable success of operational interfaces using the virtual-system principle and other techniques demonstrated in the experimental interface. The research has also suggested that certain techniques implementable on an interface could enhance retrieval effectiveness for a wide class of users by aiding them in the development of search strategies. A project bibliography and references are included. (Author/RAA)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

April 1979

Report LIDS-R-900

ED 190104

IR 008588

EXPERIMENTS AND ANALYSIS ON A COMPUTER INTERFACE TO AN INFORMATION-RETRIEVAL NETWORK

by

Richard S. Marcus

J. Francis Reintjes

This material is based upon research supported by the National Science Foundation Division of Information Science and Technology, under Grant No. NSF IST-76-82117. Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the authors, and do not necessarily reflect the views of the National Science Foundation.

LABORATORY FOR INFORMATION AND DECISION SYSTEMS

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

Cambridge, MA 02139

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Richard S. Marcus

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

ABSTRACT

Investigations have been continued and concluded on the concept of a virtual-system mode of access to a network of heterogeneous interactive retrieval systems and databases. An experimental translating computer interface, named CONIT, that enables the virtual-system mode, was developed as a research test vehicle. The interface was designed to make the basic functions of three different bibliographic retrieval systems easy to use, even by inexperienced end users, by providing a simplified common-command language coupled with extensive online instruction. Analysis of controlled experiments with end users indicates the probable success of operational interfaces using the virtual-system principle and other techniques demonstrated in the experimental interface. The research has also suggested that certain techniques implementable on an interface could enhance retrieval effectiveness for a wide class of users by aiding users in the development of search strategies.

ACKNOWLEDGMENTS

This report constitutes the final report for Grant NSF-IST76-82117 entitled "Research in the Networking of Interactive Information-Retrieval Systems". The project was supported by the Division of Information Science and Technology of the National Science Foundation. The grant period ran from May 1, 1977, to April 30, 1979.

We wish to acknowledge the cooperation of the National Library of Medicine in making their MEDLINE online-retrieval system available to us for this research.

We wish also to acknolwedge the important contributions made by the following M.I.T. graduate, undergraduate, and special students who participated in the project: Carolyn Cole, Thomas J. Deane, Esther P. Jaffe, Siu-Wah Victor Kwong, Wai-Cheong Dick Man, and Man-Kam Kenneth Yip.

CONTENTS

		Page
1.	Introduction	1
	1.1 Overview	1
	1.2 The Importance of Getting End Users Online	2
	1.3 Barriers to End-User Searching	2
	1.4 The Translating-Interface/Virtual-System Approach	3
	1.5 Outline of Current Work	5
2.	Enhanced CONIT Experimental Interface	7
	2.1 New CONIT Features	7
	2.1.1 Improvements to User Instructions	7
	2.1.2 New Functions (The True Virtual System)	8
	2.2 Example CONIT Session	8
	2.3 Summary of CONIT	17
3.	Interface Experiments	20
	3.1 Objectives	20
	3.2 The Experimental Users and Their Information Needs	20
	3.3 Experimental Procedures	22
	3.4 Experimental Results	24
	3.4.1 Quantities Measured	24
	3.4.2 Overview of Results	26
	3.4.3 User's Mastery of Commands	28
	3.4.4 Search-Strategy Considerations	33
	3.5 Evaluation of Results	52
	3.5.1 General Observations	52
	3.5.2 The Importance of Search Strategy	54
	3.5.3 Instructional Media and Learning Modes	56

CONTENTS (Continued)

4. Prospects for Computer Interfaces 59
   4.1 Interface Improvements 59
      4.1.1 Reliability 59
      4.1.2 Automation and Instructional Assistance 60
   4.2 Interface Comprehensiveness 63
      4.2.1 Data Structures 64
      4.2.2 Databases and Systems 64
   4.3 Interface Configurations 66
   4.4 Costs and Benefits 67

5. Conclusions 70

6. Project Bibliography 71

7. References 72

APPENDIX A. Explanatory Messages Requestable by User 75
B. Example CONIT Session 96
C. Offline Instruction Material 106
D. Details of Experimental Sessions 120

1. INTRODUCTION

1.1 Overview

The research reported upon here considered methods of improved utilization of existing online bibliographic information retrieval systems with heterogeneous characteristics. A primary goal of the research was to investigate the technique of interposing a computer interface between inexperienced end users and the retrieval systems. This interface would permit end users to access the systems themselves, directly, rather than through professional intermediaries, as is generally required now.

The concept investigated consists of an interface, implemented as a separate computer system which interconnects the heterogeneous systems into a network through which information is readily retrieved by hiding system differences and providing other user aids. To test the validity of this concept, there was developed an experimental translational interface that mapped many features and protocols of three bibliographic systems into a single, common or "virtual" system. In the primary operating mode of the interface, it is this virtual system with which end users interact. A "pass through" mode is also provided, however, so that a user with experience in any one system in the network can engage that system in its own language.

The three systems included in the network were ORBIT (Systems Development Corporation), DIALOG (Lockheed), and MEDLINE as implemented at two locations: the National Library of Medicine (NLM) and the State University of New York (SUNY). Collectively, these systems provide access to approximately 120 databases and some 60 million documents.

An experimental interface designated CONIT, standing for Connector for Networked Information Transfer, was implemented on the M.I.T. MULTICS computer system and evaluated in a series of controlled experiments involving six end users. A comparative study of online and offline techniques for formulating search strategies was part of the experimental work.

1.2 The Importance of Getting End Users Online

It is important to get the end user online for several reasons. Many end users prefer to do their own searching; they wish to avoid the hassle of trying to make their needs understandable to someone else. They also wish to avoid the requirement of making an appointment with an intermediary searcher which may mean an inconvenience in time, location, or both. If the end user is not present during searching, the relevance-feedback benefits of interactive searching are lost. Finally, making it easy for the end user to do his own searching has a compounded potential for reducing costs: the cost of an information specialist can be eliminated or at least greatly reduced, thus making online searching more attractive to end users; and the increased usage of retrieval thus created has potential for increasing the use factor of vendors' systems, thereby offering, at least in principle, the opportunity for further user-cost reductions through the economies of the factor of scale.

1.3 Barriers to End-User Searching

Ease of use was alleged to be an inherent feature of online retrieval systems when they were first introduced. It was proposed that, through a simple interactive dialog with the computer, any user would be able to get the information he or she needed quickly and easily. However, because early designers lacked the experience needed to appreciate the difficulty of building into interactive-computer systems the necessary human-factors features that would enable the systems to be easily used, and because of the proliferation of systems, each with its own access requirements, it is professional intermediaries acting on behalf of end users, rather than the end users themselves, who are actually retrieving the information being sought. Even the intermediaries regularly require a training period of up to a week for each new system that must be learned. Furthermore, as Benenfeld (1975)* has reported, intermediaries must spend a considerable part of each work week practicing searching in order not to "lose touch", as well as spend a number of hours per week trying to keep up with system changes and new databases Specialized characteristics of infrequently used databases are easily forgotten.

Wanger (1976) has quantified the extent to which end users (i.e., those who

---

* References are listed alphabetically by author and date in Section 6.

ultimately need and use the information retrieved) operate the systems themselves. She reports that an extensive study of retrieval-system use showed that the end user performed his own search only 7.3 percent of the time. Others who have reported on problems of end users of interactive systems include Mann (1975), Curtis (1977), and Kennedy (1975). In fact, the notion has been expressed in recent information science conference discussions that there really is no hope of bringing the end user online. Our research on this point, however, leads us to the conclusion that is is entirely feasible to do so provided certain aids to simplify use can be incorporated into the computer systems.

1.4 The Translating-Interface/Virtual-System Approach

In order to investigate means to surmount obstacles hindering convenient and effective use of the multiplicity of heterogeneous interactive bibliographic retrieval systems, the M.I.T. Laboratory for Information and Decision Systems undertook a research program to examine the feasibility of interconnecting interactive retrieval systems through computer interfaces. The computer interface is intended to achieve compatibility among systems of heterogeneous hardware and software components through use of common retrieval protocols, or by translating dissimilar protocols to a common set (see Fig. 1).

Our research program emphasized an approach in which the interface is, in effect, a common system into which, and from which, requests and results are translated automatically as they flow between user and serving systems. This approach has the virtue that a user attempting to retrieve information, when entering through the access mechanism provided by the common interface, sees a single virtual system in which all the complexities of the different retrieval systems and databases are hidden; only a single uniform system is apparent. In this way, the goal of convenient use of heterogeneous computer resources is achieved, at least for the particular application of interactive bibliographic retrieval systems.

There are four aspects of our approach which, taken together, distinguish our efforts to achieve networking: (1) we concentrated on the information

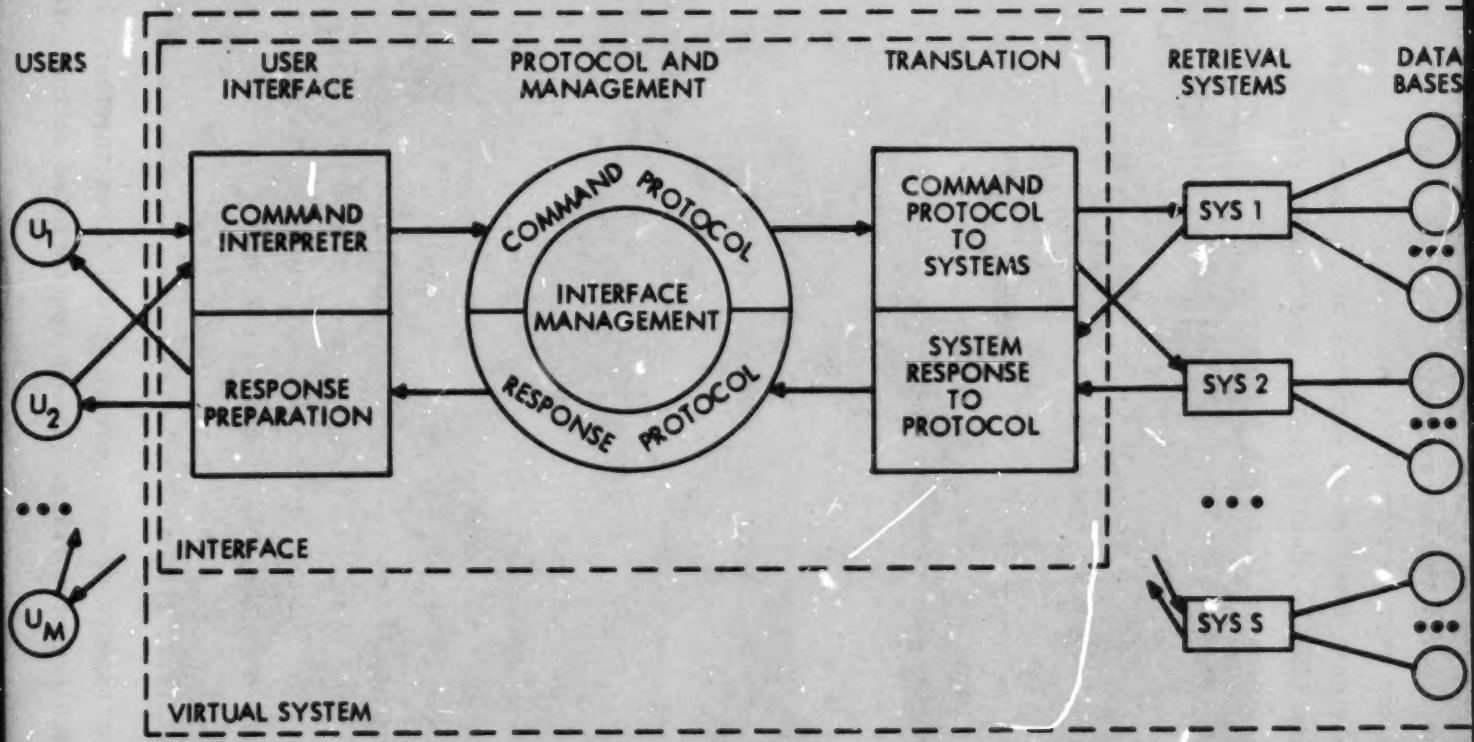

Fig. 1. Logical Diagram of Virtual-System Interface

transfer application, with particular emphasis on online bibliographic retrieval systems; (2) we utilized existing, major, stand-alone interactive systems without modification; (3) we placed emphasis on serving the ordinary end user -- that is, the user experienced neither in computer programming, general computer usage, nor in the use of interactive retrieval systems; and (4) we replaced existing, heterogeneous, often difficult-to-use computer/human interfaces with a simpler, common,easier-to-use interface.

Our research on this project has included investigations of the following related areas:

1) Networking approaches for heterogeneous systems
2) Human factors and instructional considerations for online systems
3) Retrieval functions and common languages
4) Indexing conversion and other search techniques
5) Commonality among database structures
6) Logical and software structures for networked interfaces
7) Experimental interface systems
8) User evaluation with an experimental interface
9) Cost/benefits tradeoffs

Our previous work in these areas has been detailed in reports and papers (see Section 6: Project Bibliography).

1.5 Outline of Current Work

In the remainder of this report, we describe the work that has been performed during the period of this grant, May 1977 through April 1979. In Section 2, we describe the development of an enhanced experimental interface that has been used as a vehicle for investigating the virtual-system/translating-computer-interface concept. This new development on the CONIT interface has included several aspects:

1) The addition of several retrieval functions to the virtual mode
2) Improvements to the instructional dialog
3) The incorporation of a full-fledged virtual system mode in which the user may bypass system selection and go directly to database selection

4) Information at the interface enabling users to make informed judgments on database selection

In section 3, we describe a series of experiments run with actual end users of CONIT to test the effectiveness of the various interface techniques we have employed. Effectiveness is measured in terms of the facility with which users learn the CONIT commands and the quality of the retrieved results.

Sections 4 and 5 contain an overall analysis of the interface concept and its potential in various contexts.

Sections 6 and 7 contain a project bibliography and references for this report, respectively.

The appendices contain detailed listings of instructional material and experimental usages.

2. ENHANCED CONIT EXPERIMENTAL INTERFACE

In a previous report (P 8*), we described a version of CONIT, which we could call "CONIT 2", on which experiments with end usere were performed. Results of these experiments indicated that end users could, indeed, be brought online, but that there were a number of areas in which interface capabilities needed to be enhanced before effective use by inexperienced users was possible. Since then, a new version of CONIT, which we call "CONIT 3", has been created which incorporates several enhancements to CONIT 2. In this section we first summarize these enhancements, next we give an overall description of CONIT 3 through an example session; and finally, we summarize the commands in CONIT 3.

2.1 New CONIT Features

2.1.1 Improvements to User Instructions

Experience with CONIT 2 suggested improvements to the instructional dialog which have been implemented. These included additions to, and modifications of, messages given by CONIT. Some of these messages are explicitly requested by the user by the EXPLAIN (HELP) command, and others are given by CONIT as part of the regular instruction in a given context [e.g., what to do after the response to the FIND (search) command is given]. Attention was given to providing online instruction when it would be most useful, and inhibiting overly repetitious messages. Thus, for example, a special message was added after the second search reminding the user of the COMBINE command. Another special message after the first (but no other) user error in giving a command name reminded the user about how to correct typing errors. The full list of requestable (EXPLAIN) messages for CONIT 3 is given in Appendix A. A partial accounting of the context dependent messages is given explicitly in the sample user dialog given in Section 2.2

As will be explained in detail in Section 3, various instructional materials, in addition to the online instruction, were prepared in printed form for offline use. These included a printed version of the EXPLAIN messages (see Appendix A) plus special instructions for the development

---

* "P-number" references refer to documents in the Project Bibliography (Section 6).

of online searching strategies and the selection of databases .

2.1.2 New Functions (The True Virtual System)

CONIT 2 permitted most of the basic retrieval functions to be requested in the common CONIT language. In CONIT 3, a number of additional functions were made expressible in the common langage. These additional functions include: (1) author-name searching (as well as subject-term searching); (2) the ability to request offline output from the remote host retrieval system; and (3) the ability to get information about a database and/or select a database at any point in the online session, regardless of what, if any, retrieval system is currently connected.

This third added capability is extremely important, because it realizes the virtual-system concept. By that, we mean it is no longer necessary for the user to be concerned about the different retrieval systems, as such. She or he need be concerned only about the databases to be searched.

This capability was provided by having information about each database in the CONIT interface. A user, then, can find out about available databases directly from CONIT. When a user selects a database for searching, CONIT checks to see what systems have that database. If the database is available on the currently connected system, than a simple database selection command is issued in the command language of the current system. Otherwise, a system that does have the given database is connected to automatically by CONIT (after automatically disconnecting the current retrieval system, if any). Thus, the user can ignore the issue of different retrieval systems either with respect to which systems have which databases, or which system (if any) is currently connected. Of course, the differences among system command languages had already been submerged in CONIT 2. System connections for CONIT 3 are shown in Fig. 2.

2.2 Example of CONIT Session

In order to understand the experiments and analyses in subsequent sections of this report, it is helpful to have an appreciation of the features and capabilities of the CONIT interface system. Perhaps the

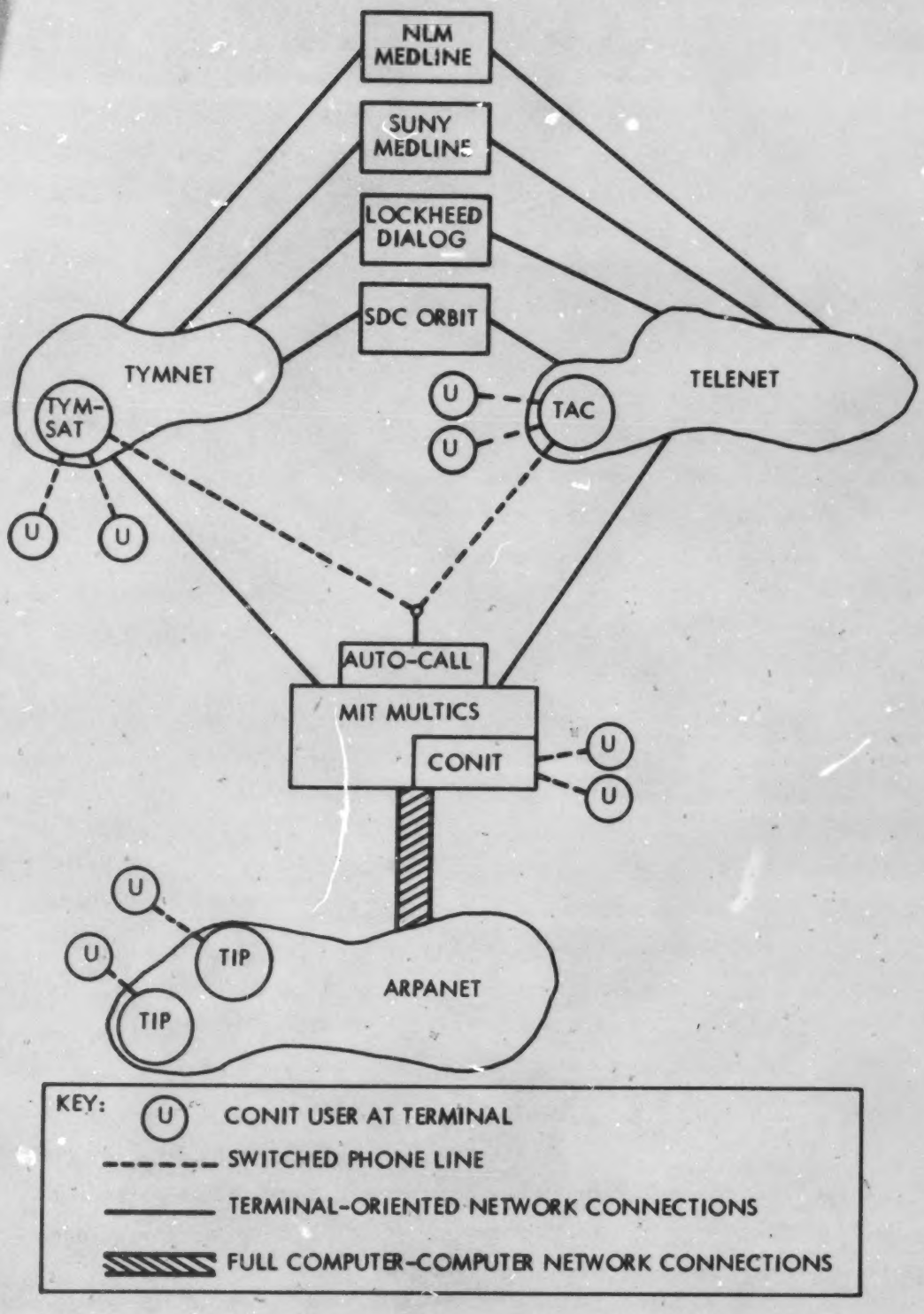

Fig. 2. CONIT 3 Experimental Network

best way to gain this appreciation is to review the operation of CONIT for a typical online session. For this purpose, we have chosen the online session of one of the CONIT users in our last round of experiments. Excerpts of a slightly modified version of this session are reproduced in Appendix B. The modifications include a few additions to the actual session, in order to illustrate certain features. The user commands are labeled U1, U2, ... The CONIT system responses are labeled C1,C2, ... A reproduction of one page of the original typescript is shown in Figure 3.

Prior to the start of the session itself, the CONIT interface system has been logged in and set up at a terminal connected to the MULTICS computer. The user is told to start his session by typing the word "start" followed by a carriage return on the terminal (U1). CONIT's response (C1) is a message welcoming the user and giving him a simplified method for correcting typing errors and explaining how to get instructional help. While help is suggested, the user is given the option of going off and doing what he wants (a general principle in CONIT).

Two aspects of the user/system interactive dialog are introduced in this first message: (1) the user should wait for the user cue which indicates that it is his turn to "talk", and (2) the user indicates the completion of his command by a carriage return. Two other points are illustrated by message C1: (1) suggestions on what to do next are given and highlighted by special format (separate line with five space indentation -- contrasted with three-space indentation for paragraphing); and (2) explanatory information is given in "chunks" with the size of the chunk being related to the context (messages early in the session are generally shorter and cover fewer facts than later ones).

The user follows the system advice and asks for help (U2). In its response (C2), CONIT introduces the EXPLAIN command. In doing so, CONIT also introduces through example the idea of the command-name/argument format for the user command langauge, and the idea that abbreviations are permissible for shortening user commands.

```
TO SELECT ANY OF THESE TERMS FOR SEARCHING TYPE: E FIND INDEX

1030.4
*****USER::
FIND MODEL

++++CONIT:
NO RESPONSE RECEIVED FROM LAST TRANSMISSION.
SHOULD CONIT WAIT FOR A RESPONSE ANY LONGER? (ANSWER YES OR NO)
   RETRIEVAL SYSTEMS MAY BE SLOW IN RESPONDING AT TIMES.  YOU SHOULD
ANSWER YES TO THE ABOVE QUESTION ABOUT 3 SUCCESSIVE TIMES (UNTIL ABOUT
1 MINUTE HAS ELAPSED) AFTER WHICH TIME YOU MAY ASSUME EITHER THE RETRIEVAL
SYSTEM OR CONIT OR THE NETWORK CONNECTION BETWEEN THE TWO
IS NOT WORKING PROPERLY.  AT THAT POINT YOU SHOULD ANSWER NO TO THE
QUESTION.  THEN TRY YOUR REQUEST AGAIN.  IF YOU STILL GET NO RESPONSE,
DISCONNECT FROM THE RETRIEVAL SYSTEM (TYPE DIS) AND TRY TO PICK IT AGAIN
(POSSIBLY USING A DIFFERENT NETWORK) OR PICK A DIFFERENT SYSTEM.
*****USER::
YES
NO RESPONSE RECEIVED FROM LAST TRANSMISSION.
SHOULD CONIT WAIT FOR A RESPONSE ANY LONGER? (ANSWER YES OR NO)
*****USER::
YES
FROM HOST:
YOUR SEARCH RESULTED IN SET 1 WHICH CONTAINS 10415 DOCUMENTS.
TO SEE REFERENCES TO THE FIRST FEW DOCUMENTS TYPE: SHOW

1034.5
*****USER::
FIND WORLD

++++CONIT:
NO RESPONSE RECEIVED FROM LAST TRANSMISSION.
SHOULD CONIT WAIT FOR A RESPONSE ANY LONGER? (ANSWER YES OR NO)
*****USER::
YES
NO RESPONSE RECEIVED FROM LAST TRANSMISSION.
SHOULD CONIT WAIT FOR A RESPONSE ANY LONGER? (ANSWER YES OR NO)
*****USER::
YES
FROM HOST:
YOUR SEARCH RESULTED IN SET 2 WHICH CONTAINS 6904 DOCUMENTS.
TO SEE REFERENCES TO THE FIRST FEW DOCUMENTS TYPE: SHOW
FOR AN EXPLANATION OF HOW TO COMBINE SETS TYPE: E COMBINE

1036.7
*****USER::
COMBINE SET 1 AND SET 2

++++CONIT:
YOUR SEARCH RESULTED IN SET 3 WHICH CONTAINS 113 DOCUMENTS.
TO SEE REFERENCES TO THE FIRST FEW DOCUMENTS TYPE: SHOW
TO REVIEW YOUR SEARCH SETS TYPE: SHOW REVIEW

1037.9
*****USER::
SHOW

++++CONIT:

2/3/1
785689
  HOUSING REQUIREMENTS AND NATIONAL RESOURCES - IMPLICATIONS OF UN
ORLD MODEL (EN)
  LAKSHMANAN TR; CHATTERJEE L; ROY P
  JOHNS HOPKINS UNIV,CTR METROPOLITAN PLANNING &
ES/BALTIMORE/MD/21218; JOHNS HOPKINS UNIV,DEPT GEOG & ENVIRONM
```

Fig. 3. Page of Typescript for Experimental User (with annotations; reduced 30%)

In C3, the user sees a list of the basic commands and a short explanation of each. This amount of detail was designed to be sufficient to give the user an overview of the interface system capabilities and a guideline for future instruction without overburdening him/her with details. This goal is aided by having relatively few different basic commands. C3 also gives additional instruction on the basic procedures for the interactive dialog. Much of this instruction has been seen by the user before; the technique of repeated instruction is a purposeful approach in order to achieve learning reinforcement. In C3 the concept of a hierarchical structure to instruction is brought forward. The simple, general explanation is given first; more detailed explanations are available, at user request or system discretion, to elaborate on various aspects of the initial explanation. The inexperienced user is prompted to get an explanation of the PICK command next; this follows the principle of giving, or suggesting, instruction on a system feature at the first point in the session when the user is likely to need the feature.

In C4, CONIT explains how the PICK command is used to pick databases for searching. In accordance with the virtual-system approach, the user is not required to consider what retrieval systems are appropriate. However, information is given on how the explanation for system specification, which a sophisticated user might want, can be requested.

The response to the SHOW DATA command (C5) lists the seven subject areas chosen as a high-level classification scheme for the databases to which CONIT has access. Again, following the hierarchical approach presented by CONIT, the user* specifies one of these areas (Area 5) for detailed database listing in the command SHOW DATA 5 (U6).

For each database, CONIT gives a common CONIT name and number, the "standard" name, a short explanation of the database, and an indication of which retrieval systems have the database. The CONIT name is intended to be more indicative of the subject content of the database than the "standard" name which is often a meaningless character string to the uninitiated (note that different systems sometimes give different names to the same database). Any of these names is allowed for use in the PICK command, as

* The actual experimental user on which this sample session is based did not issue commands U5 or U6, since he was in the group of users who were presented offline printed explanations of databases (as explained below in Section 3).

well as the CONIT number, which provides a short form that incorporates the CONIT classification scheme.

Generally, the short explanation of a database is sufficient information upon which a user can decide whether to search the database. However, additional information is available through the EXPLAIN DATA name command. (Name is the name, or number, for the database on which information is desired.) By this command, the information about a database available online from a retrieval system is requested of that system. If the retrieval system which has the information is not currently connected CONIT connects to it (see below for discussion of automatic connection protocols). The information itself is obtained by CONIT's sending the FILEn and ?FIELDn commands to DIALOG (n = the DIALOG number of the database) or the "EXPLAIN name" command to the ELHILL* or ORBIT systems (name = database name as known by the host retrieval system).

Next, the user picks the SOCIAL SCIENCES CITATION INDEX database (CONIT name SOCSCI, CONIT number 51), with the command PICK 51. CONIT recognizes that the database is available only on the DIALOG system (at the time of this experiment) and sets about connecting to DIALOG. Note in this virtual system mode of operation the user need not be concerned about where the database actually is; CONIT will find it for him. If the database is available on more than one system, CONIT will choose according to the following algorithm:

1) If there is a currently connected system and the database is available on it, that one is chosen;
2) Otherwise, the "normally preferred" system is chosen. (Preference is now preset, although it could be dynamic, and is based on such criteria as which system is more readily available or which database implementation is more complete or less costly. For example, SUNY/MEDLINE is preferred over NLM/MEDLINE because it is usually less busy, and for that reason, is the system users are urged to use by the National Library of Medicine for accessing the MEDLINE, SDILINE, or MESH vocabulary

* 'ELHILL' is the official name for the National Library of Medicine's retrieval system, which is often known by its largest database, MEDLINE.

databases.)

When a database is picked, and no system is currently connected, as is the case in this example, CONIT must request the MULTICS system to establish an appropriate physical network connection. This requires the placing of a telephone call which is accomplished by the automatic dialing equipment. The physical connection takes about 30 seconds during which time the messages given in the first nine lines following the '++++CONIT:' header in C7 are presented to the user. These messages serve four purposes:

1) Feedback concerning CONIT's understanding of which database was picked is given the user for checking purposes.
2) The identification of the interconnecting network* (here TYMNET) and the system chosen is given; this information is not necessary, but may be useful -- for example, in case system or network is unavailable.
3) Additional instructional information is given to the user concerning his interactive dialog with the interface.
4) The user is kept informed of what is going on, and not left to wonder, because he sees nothing happen, if something has gone wrong.

The fourth purpose provides the main reason for giving the remaining parts of the response C7. Each of these lines is given after some corroborated response is obtained in the sequence of: (1) placing the telephone call; (2) identifying the terminal properly to the network and inhibiting network echoing of characters; (3) establishing the connection to the host retrieval system; (4) logging into the retrieval system, with appropriate identification and/or password strings; (5) in some systems, establishing the abbreviated version of the dialog (between CONIT and the retrieval system), and, finally (6) connecting to the desired database. The entire process is automatic, and while messages are given the user to keep him assured that progress is continuing, he need not be concerned with any of the many details of the connection process. One aspect of the host messages of the connection process is fed back to the

* The network is chosen by CONIT if the user does not specify. The default network was TYMNET here; at other times it has been TELENET. The preset choice of network is based primarily on which network was easier to make local connections to in the recent past.

user for his possible interest: any broadcast news message. No such message appeared in this login, but such news was fed back in the login to ORBIT shown in C56 below.

The explanation of the FIND command (C8) gives the user an exposition of the keyword/stem, free-vocabulary-based initial search strategy recommended by CONIT. The user then requested explanations (C9-C14) of additional information on searching, as recommended by CONIT.

At this point in the session, a communications failure occurred. The user picked the database again, and communications were re-established. The user, who was searching on the subject "world economic and political models" then followed one of the previous CONIT suggestions and used the SHOW INDEX command to browse the index of the SOCSCI database for terms alphabetically near two words related to his search: "model" and "world". The SHOW INDEX command is translated as the EXPAND command in DIALOG and the NEIGHBOR command in ELHILL and ORBIT.

The user follows the CONIT suggestion to search on keyword stems with his first search: FIND MODEL. Following CONIT's searching principles of searching all possible indexes on stemmed forms, this is translated as SELECT MODEL? -- i.e., do a truncated search on the given word in the basic index. In ELHILL and ORBIT, this would be translated as "FIND ALL MODEL: These latter two systems do specifically allow searching all indexes, in contrast to DIALOG, which requires a different kind of search request for each non-basic index.

The host response from this search takes longer than a preset figure (here set at five seconds). Whenever long delays like this occur, CONIT gives the user the option to wait longer for the response, or give up and go on to something else. This procedure is another manifestation of the doctrine of avoiding excessive time during which the user might be unsure of what, if anything, is happening or what he should do in the absence of response. Of course, this procedure also handles the case where, due to some unforeseen occurrence, the host system never will respond. We may note that the detailed explanation on what to do in this delay situation is presented to the user only in the first instance;

in later instances, it is assumed that the user knows what to do.

The power of the truncated or stemmed search is demonstrated by the fact that the results for set 1 and set 2 show significantly higher retrieval counts than would untruncated searches.

Note the special message at the end of the response to the second search, suggesting the combine operation; this follows the principle of providing suggestions when most likely to be needed. The user here had already seen the explanation on the COMBINE command and so goes ahead to perform it.

The user next issues the SHOW command to see catalog output for retrieved documents from the current set. After getting more detailed information on how to get more particular output information, the user requests all the information available on document 4. He then requests all on all documents in the current set be output offline (at the remote host site) and mailed to M.I.T.

Just a few interactions selected from the remainder of the session will be chosen to illustrate some other system features. In U33, the user has made a syntactic error (left out space after COMBINE command); he realizes this and strikes the BREAK key to "kill" the line up to that point and return to CONIT command level in order to resubmit command in correct form. In U53, the user employs the Boolean AND NOT operator in a COMBINE command to eliminate documents already retrieved. In U56, he decides to search a new database: SSIE (the Smithsonian Science Information Exchange database containing research project summaries). At the time of the experiment, this database was available through CONIT only on the ORBIT system. CONIT recognizes this fact and logs in to ORBIT after logging off of DIALOG.

The user employs the same CONIT commands to redo his search in SSIE; of course, different translations are involved. In U84, the SHOW REVIEW command is given to request a listing of all searches done on SSIE during the session until that time. In U90 the session is terminated

with the STOP command; CONIT automatically logs off of the connected retrieval system before terminating the session.

2.3 Summary of CONIT Commands

The functional capabilities of CONIT are summarized in the following list of commands:*

2.3.1 Instruction Requesting

EXPLAIN concept

An explanation of the concept or item designated concept is given

HELP = EXPLAIN EXPLAIN (Get-started message given)

2.3.2 Database Information

Show DATA

Lists seven areas of data bases

SHOW DATA n

Lists databases in area n

EXPLAIN DATA database [system]

Get the explanation of database database from system system

2.3.3 Database Selection

PICK database [system] [network]

Connect to database database on system system through network network. First log off currently connected system and select default system and database, as required.

2.3.4 Search and Index Operations

FIND subject

Do a truncated search in the basic index (and in other indexes, if possible) on the subject term(s) subject.

---

* In the explanation of commands, we use underlining in examples of language construction to indicate variable elements and bracketing to indicate optional arguments.

FIND AUTHOR author-name

Do a truncated search in the author index under the name author-name.

SHOW INDEX term

Perform the browsing command to display terms alphabetically near term from the basic index of the currently connected database.

SHOW REVIEW

All searches in the currently selected database are reviewed.

COMBINE SET i bool SET k

The sets designated i and k are combined according to the Boolean operator bool, where bool = "and", "or", or "and not".

2.3.5 Outputting Document Information

SHOW [DOCUMENTS i - k] [SET s] [TYPE] [OFF]

Information of the type type for documents numbered i through k of set s is given. The types allowable are "title", "abstract","citation (title, author, and source)", and "all"; the default type is "citation". The default documents are the first five. If k is missing, information on the single document i is output. If OFF is present, the information is output offline; otherwise, online display is assumed. The order in which the arguments appear is optional.

2.3.6 Miscellaneous Commands

SEND command

The command command is sent directly to the connected remote host retrieval system without translation.

START

Begin, or restart, session.

STOP

End session (disconnect any connected systems first)

DISCONNECT

Disconnect connected systems (but do not end session)

COMMENT comment

The user's comment comment is saved by CONIT for future review by the research project investigators.

SHOW STATUS

The names of the currently connected database, system, and network are displayed.

SHOW NEWS

The current news items from the connected retrieval systems are displayed to the user at the terminal.

In addition to the commands meant for regular use by ordinary end users, there are a number of commands designed for use by the CONIT system implementers and operators. These commands permit the following operations: selecting a rule table to guide the session and translations; making modifications to the current rule table; and putting the CONIT system in a mode to accommodate debugging operations.

2.3.7 Interrupting and Editing

The BREAK key can be used to interrupt action of CONIT or the remote system. If CONIT action is being interrupted, control is returned immediately to the user for further commands. If remote-system action is being interrupted, then CONIT sends a simulated BREAK (series of nulls) to that system (if it accepts breaks); in any case, CONIT then waits for a signal from the remote system signifying that it is ready for additional commands before returning a user cue to the CONIT user.

The MULTICS character-cancel code (#) and line-cancel code (@) are used. In addition, the BREAK key can be used to kill the line and get a new user cue.

3. INTERFACE EXPERIMENTS

3.1 Objectives

The broad objective of the experimental program was to evaluate the CONIT interface in terms of its ability to allow end users to satisfy their informational needs by accessing a network of heterogeneous information systems through the interface. Several measures of effectiveness were considered: quantity of information recovered; usefulness of recovered information; effectiveness of individual searches (as measured by precision) and of the overall session (as measured by recall); the time required to arrive at various stages of the search; and general user satisfaction with the interface.

A second objective was to make a comparative analysis of an online and offline approach to search-strategy formulation. As discussed in detail below, we have come to the conclusion that an end user's ability to formulate a good search strategy is crucial to his success in extracting useful information from the network. Trained information specialists have considerable skill in arriving at a good strategy; how to substitute for this skill when an inexperienced user is online was one segment of our investigations.

3.2 The Experimental Users and Their Information Needs

We sought typical end users to serve as experimental users (EU's) of the CONIT interface system. Such users were solicited through notices posted at locations scattered around the M.I.T. campus. These notices offered free information from computer databases in return for participation in our experiments. Responders to the offer were interviewed to determine their information needs and backgrounds, especially with respect to computer experience. Basically, the six EU's were the first responders who could be fitted in with the experimental schedule. Several responders were turned away, either because they were already proficient searchers in one of more computer retrieval systems, and/or they did not

have an immediate, bona fide need in terms of a well defined topic of current personal interest.

The general search topics and professional backgrounds of the six chosen to be EU's are given in Table 1. All six EU's had used computers online to a limited extent, but not for information retrieval. However, four of the six had observed others do online bibliographic searching.

Table 1. General Search Topics of the Six EU's

EU No.	Professional Level	General Search Topic
1	Graduate student in Library Science; part-time secretary	The interface between municipal libraries and social service agencies
2	Engineer; part-time graduate student in Aeronautics	Airplane wing section design
3	Graduate student in Mechanical Engineering and in Political Science	World models in economics and politics
4	Graduate student in Physics	Cross sections of helium ions under electron bombardment
5	Research Staff member; Post-doctoral Fellow	Measurements of two-ear phenomena in persons with a hearing loss
6	Graduate student with background in Chemistry, Library Science and Social Science	Oral examinations as a testing technique and curriculum design in medical education

None of the EU's had been using computers regularly in the recent past. One of them was an excellent typist, four were fair, and the remaining one classified himself as a "hunt and peck" type. All of the EU's were in an academic and/or research environment. Five of the six were currently studying at the graduate level: two full-time M.I.T. doctoral candidates, one special (part-time) M.I.T. graduate student, and two Master's degree candidates at neighboring colleges. The sixth EU was a post-doctoral Fellow.

The search topics covered a wide variety of subject areas, including three in social science areas and three in physical, medical, or engineering sciences.

3.3 Experimental Procedures

All experiments were carefully controlled and monitored. To the extent possible, the same set of procedures was used for each EU and the same set of data points was recorded. Three EU's (EU1, EU2, and EU3), identified as "Group A", began the development of their search strategies offline, in accordance with a set of printed instructions given to them at the outset of the session before they engaged the computer terminal. These instructions concern database selection and search-strategy formulation; they are reproduced in Appendix C.

The other three EU's (EU4, EU5 and EU6), identified as "Group B", did not get the offline instructions; they were assisted solely by the online instructions given through the terminal. In other words, all EU's were given online instruction through CONIT, while just the first three had additional offline instruction.

At the beginning of each experimental session, an experimental supervisor (ES) briefed the EU concerning the nature of the experiment. In order to keep this briefing as nearly uniform as possible, a common briefing statement (see Appendix C) was read to the EU. The oral briefing took from 3.5 minutes to 6.0 minutes, and averaged 4.6 minutes in length for the six EU's.

Following the briefing, the Group A EU's were given the offline instructions. These users spent between 20 and 36 minutes with these materials, as shown in Table 2. The experimental supervisor answered the EU's questions after the debriefing and during their use of the offline instructions. The online session itself — i.e., the EU's interaction with CONIT at the computer terminal — began immediately after the briefing for the Group B EU's and immediately after the offline preparation mentioned above for the Group A EU's. The experimental supervisor remained in the room where the EU worked at the terminal. The purpose

of this arrangement was to provide unobtrusive monitoring of the session and of the EU (beyond what was possible from the full record of user commands and CONIT responses, as maintained in a computer "audit" file). The supervisor sat at a terminal located at the opposite end of the room from the one at which the EU worked. At the terminal, the supervisor feigned work unrelated to the EU, but actually used the terminal to monitor the CONIT/EU interactions. The supervisor could also observe visually any activity of the EU not recorded on the computer. The EU himself was asked to make written or mental notes of particular problems or other reactions he might have, so as to be able to relate his/her experiences more completely to the supervisor at a post-session debriefing. The supervisor did not prompt the EU or otherwise interfere with the course of the interaction unless computer system bugs or computer communications problems appeared to hinder the continuation of the session; a few such problems did crop up, as explained below.

The online session was terminated when the EU issued the STOP command. The primary motivation for terminating the session appeared to be the EU's feeling that he was at the point of diminishing returns, as far as getting useful information was concerned. At the end of the online session, the supervisor held a debriefing conference with the EU, as a means of getting his reactions while they were fresh in his mind. This debriefing included a detailed review of the session interactions using the terminal typescript and any note of the EU or supervisor as guides. The review emphasized: 1) any problems encountered by the EU; (2) the rationale for EU search strategy formulation; and (3) relevance judgments by the EU on retrieved documents. The EU's were asked to judge relevance on a four-point scale: high, medium, low, and none.

Based on the relevance judgments given by the EUs during the experimental sessions, we subsequently performed an analysis of search strategies that might improve on the searches of the EU's. Searches based on this analysis, both in databases picked by the EU and in others, were then performed. Catalog output from selected documents retrieved from these

searches, along with the offline output requested during the online experimental session, was then presented to the EU for further relevance judgments. In a couple of cases, a second round of analyst searches and EU relevance judgments was carried out before a final analysis of the experiment was completed.

3.4 Experimental Results

3.4.1 Quantities Measured

Statistical data derived from the various stages of the experimental analyses are summarized in Table 2. The measurements for the individual EU's are given along with averages for each quantity measured. Averages are shown for: (1) Group A, the offline-instruction users (EU1-3); (2) Group B, the users having online instruction only (EU4-6); and (3) all six EU's (EU1-6) taken as a single group.

The quantities measured include the times required to reach various critical junctures of the session and numbers related to how busy the EU was and how many useful documents were retrieved, compared to the number potentially retrievable in the databases. The times measured in (2a) through (2d) refer to the time from the beginning of the online session at the terminal. The time-to-first-PICK (2a) is how long it took the EU to pick a database in which to search. The time-to-first-FIND (2b) is the time until the first search command was issued. The time-to-first-SHOW (2c) is the time until the first output on retrieved document references was requested. The time-to-first-useful-reference is how long it took before the EU saw a document reference that was useful to his or her problem.

The number of commands issued online by the EU (4a) and the resulting number of online typescript pages (4b) -- each page is about 60 lines, some of which are blank -- (see Fig. 3 for typical page) are measurements of "busyness". These figures are normalized for length of time at terminal in (5a) and (5b).

The number of document references requested by the EU (8) is broken down to those shown online (8a) and those requested for offline output (8b). Similarly, the number of documents judged useful by the EU (9) -- based on

Table 2. Results of User Experiments

Quantity Measured	Individual Users						Averages		
	EU1	EU2	EU3	EU4	EU5	EU6	EU1-3	EU4-6	EU1-6
1. Time (min) for offline instruction	20	23	36	-	-	-	26	0	–
2. Time (min) online until:									
2a. first PICK	5	9	7	6	23	6	7	12	9
2b. first FIND	8	14	34	15	32	11	18	19	19
2c. first SHOW (Documents)	20	19	41	15	41	15	27	24	25
2d. first useful ref.	30	29	41	47	50	15	33	37	35
3. Total online time (min)	50	57	110	69	124	108	72	100	86
4. Amount of interaction									
4a. commands issued	63	38	90	72	114	81	64	89	77
4b. pages of typescript	14	11	24	19	30	27	16	25	21
5. Rate of interaction									
5a. Commands/minute	1.3	0.7	0.8	1.0	0.9	0.8	0.9	0.9	0.9
5b. Pages/hour	17	12	13	17	15	15	14	16	15
6. Total session time (Min)	70	70	146	69	124	108	98	100	99
7. Number of:									
7a. Databases searched	2	1	2	2	2	2	1.7	2	1.8
7b. Retr. systems used	2	1	2	1	2	2	1.7	1.7	1.7
8. Number doc. references:									
8a. shown online	20	7	68	29	60	38	32	42	37
8b. printed offline	40	50	319	0	0	0	136	0	68
9. Number useful refs:									
9a. seen online	4	3	34	1	31	8	14	13	14
9b. total found by user	19	30	171	1	31	8	77	13	45
10. Recall base, est.	960	1000	730	140	1200	670	890	670	780
11. Recall	.02	.03	.23	.01	.03	.01	.09	.02	.05

the particular catalog output seen -- is indicated for online retrieval (9a) separately from the total found (9b). Precision figures are implicit in these statistics, although the precision so derived is for selected portions of selected searches.

The recall base was estimated on the basis of relevance judgments made by the EU on samples of documents retrieved in post-session analyst searches. For purpose of this summary, a document was considerd "relevant" if it was marked as being of "high" or "moderate" relevance by the EU; "low"-relevance documents are lumped with "no"-relevance documents as being nonrelevant. For the purposes of the six EU's, relevance was closely related to usefulness, although we shall qualify this point later. The recall base was derived by a simple extrapolation on the samples shown to the EU: as will be discussed below, this is simply a lower bound on the actual recall base that would be found through fully exhaustive search methods.

3.4.2 Overview of Results

The results generally show that the EU's were able to use CONIT and find a number of relevant documents on their own. However, because particular circumstances affect the results so strongly, we need to consider more details concerning the individual sessions in order to properly assess the degree of effectiveness of various aspects of the interface. A summary of pertinent session details is given in Appendix D.

In attempting to evalute the effectiveness of the experimental CONIT system -- and thereby, the potential effectiveness of similar interface systems -- we might start by considering one measured quantity in Table 2: the online time until the first useful reference was retrieved (2d). The average of this measurement over all six EU's was 35 minutes. This figure can be judged in either a negative or a positive light. On the one hand, if the EU had worked with an information specialist intermediary who was experienced in the several systems involved, the same results could undoubtedly have been achieved more quickly.

On the other hand, the experimental results suggest that inexperienced

users can, with the aid of interface techniques, find information from heterogeneous systems and databases in reasonable periods of time. Note that what is included within this time period is learning how to interact with CONIT at the terminal, and specifically, learning and using commands to get explanations, pick databases, perform searches, and display information on retrieved documents. In addition, for the EU's of Group B (and partially for those of Group A), information had to be absorbed regarding the nature of particular databases and the techniques for search-strategy formulation.

In making more detailed evaluations of the interface and its potential, it is worthwhile to consider two related but separable aspects of the EU's use of the system: (1) the mastery of the CONIT command language and modes of interaction, and (2) the adequacy of the development of search strategies. However, before looking at these two aspects in detail, we first take up the question of supervisor's assistance.

Since there were a number of instances of supervisor interaction with the EU during the online session (see Appendix D), it is appropriate to consider the extent to which the results were altered by this deviation from the desired goal of no human interference. Our opinion is that while some documents might not have been retrieved, or might have taken longer to retrieve, without supervisor aid, the basic conclusions about potential interface effectiveness will not be affected.

The justification for this opinion is based on the observation that there were three main kinds of situations in which the supervisor intervened. In the first type of situation, there was a CONIT system bug that seriously interfered with the experimental session. Two such situations arose: (1) inadequate handling of "TIME OVERFLOW" message from ORBIT for EU3 (Section D3.2) and (2) incorrect handling of the BREAK request for EU4 (Section D3.5). In the second type of situation, a clear inadequacy in CONIT operation was perceived. This might be termed a kind of <u>design</u> bug as opposed to the implementation bug of the first situation. Again, there

were two such situations: (1) confusing message when delay occurs in the host system response (see Section D1.2); and (2) the incorrect handling of a search on terms found in a SHOW INDEX response (see Section D4.9).

In the third type of situation, the supervisor simply decided to terminate the session before the EU was fully prepared to do so on his own. This happened with EU2 when the supervisor advised him on how to get off-line output in order to finish the session.

In other cases of CONIT, communications, or user problems and questions, the supervisor said nothing or merely urged the user to attempt to handle the problem himself. In no case did the supervisor interfere with search strategy formulation or otherwise push the EU to use a particular command or approach.

In a working interface environment, the first two types of situations would be ironed out after a suitable debugging period. In fact, all of these problems uncovered in the experiments were soon corrected in CONIT. The problem with EU2, on the other hand, has been accepted as a partial failure of the CONIT system used by the EU's, although in fact EU2 was satisfied with the results he had obtained online.

3.4.3 User's Mastery of Commands

In evaluating the EU's mastery of CONIT commands, we may once more look at both negative and positive aspects. On the negative side, as is detailed in Appendix D, four of the EU's each made at least a few mistakes in using CONIT commands. On the positive side, each EU did, with the help of the system, successfully use, either at first or eventually, each basic command. Furthermore, with the exception of EU2, who blundered into the use of several commands without really understanding them, each EU arrived at a reasonable understanding of all the basic commands, at least in their simple forms.

It is instructive to consider the mix of commands used by the EU's. Table 3 shows the eight most commonly used commands issued by the EU's. It may be noted that, while all EU's were in a learning mode, well over half of the commands were "working" commands, i.e., commands requesting searches, search terms, or document output -- as opposed to commands simply asking for explanations of system use or requesting continued waiting for a host response (YES).

Table 3. Most Commonly Used Commands

Command Name	Average Number of Usages, per User
FIND	13.3
YES (response to the wait question)	12.8
EXPLAIN	10.3
SHOW (Documents)	9.2
COMBINE	6.9
SHOW INDEX	3.8
PICK	2.8
BREAK (key hit)	2.7

While all the basic commands were used by each EU, there were a number of specialized commands and explanations which could have been more helpful if used by more of the EU's. The SHOW REVIEW command was used by only two EU's; others would have been able to keep track of previous searches more easily if they had used it. None of the EU's used SHOW TITLE, although that could have speeded up browsing through document output. EU1 did not use SHOW INDEX or SHOW ALL -- two functions that could have helped her search strategy formultation.

Three EU's did not avail themselves of the opportunity to get offline output. The explanations for developing search strategies -- beyond the initial E FIND -- received scattered usage. Table 4 shows the number of EU's who requested these explanations.

Table 4. Use of Search Explanations

EXPLANATION	NUMBER OF EU'S REQUESTING
E FIND	5
E FIND AUTHOR	3
E FIND BETTER	3
E FIND MORE	1
E COMBINE	4

As will be discussed below under search strategy analysis, the effectiveness of use is related to the thoroughness with which the EU's sought and read the explanatory information available online.

To delve further into the effectiveness with which EU's used CONIT, we may consider the time required to perform particular tasks. One approach to this analysis is to look at the time measurements in Table 2, and in particular, at time differences. In recording the "time to first PICK", we are measuring how long it took the EU to get, read, and digest the initial online explanations about system use and commands -- particularly, EXPLAIN and PICK commands -- as well as getting sufficient information about databases in order to select one. The average for all six EU's was nine minutes. The Group A EU's (with offline instruction) did considerably better (seven minutes) compared with the Group B (no offline instruction) EU's (12 minutes). It is tempting to attribute this difference to the availability of database information to Group A prior to the online session. However, a detailed review of the individual EU session, which we shall now give, casts some doubt on this simple interpretation.

It does appear true that each of the EU's was able to get and read the initial explanations with reasonable dispatch. Also, none of the Group A EU's used any SHOW DATA command to get online information about databases, whereas the Group B EU's used an average of three such commands each. However, if we look at the time to pick for the individual EU's, we see that it is only the time for EU5 (23 minutes) which is significantly different from that of the other EU's, and is the cause for the large difference between the two groups. When the session for EU5 is reviewed

(see Appendix D), it is seen that a major reason for the long time-to-pick by the EU was his decision to get additional information about one particular database and the difficulty in getting that information because of telecommunications problems and a misreading of the database number. Except for these problems, the EU5 time-to-pick would have been about 10 minutes, and Group B's average would have been close to Group A's seven-minute average.

On the other hand, the one EU who was most helped by the offline database information was EU2. Unlike most of the others, he had little a priori knowledge of any online database, and he used the alphabetical index to databases by professional fields (see Appendix C) to learn that aeronautical engineering was covered by the COMPENDEX (ENGINEERING INDEX) database. Yet it was EU2 who spent the second most time to pick (nine minutes) of any of the EU's. The explanation for this anomaly is merely that EU2 spent more time on the early command explanations -- due to his many typing errors and other confusions.

An analysis of the utility of offline database instruction suggests that this method of presentation could save two to three minutes of online instruction for users like our EU's. While the overall system cost might be somewhat lower with offline instruction, the total user's time might be expected to be about the same, assuming similar information available from both media. Additional comparison of offline and online modes is given below in Section 3.5.3.

This analysis illustrates the need to look behind the simple averages to the individual circumstances in order to interpret the statistics properly. In particular, considerations of individual background, needs, and errors, as well as system problems, need to be taken into account. This is especially true in the situation where relatively few individuals are being grouped in each statistical group.

The time from the first PICK to the first FIND command (2b - 2a) includes the time to make the connection to the database and get online instructions on searching, read the instructions, and prepare the first

search. The average time for the six EU's was 10 minutes. It might be expected that Group A, who had the pre-session offline instruction, would be faster. However, Group B is actually faster on average; seven minutes compared with 11 minutes. Here again, we must look behind the averages. The longer average for Group A can be traced to EU3, who took 27 minutes in this period. When we analyze the reason for this length of time, we find that EU3 was the only EU who requested all of the online search instructions, despite having also seen the offline instructions. EU3 was meticulous in his utilization of both offline and online instructions; in this respect, it may be noted that he also spent the most time (36 minutes compared to an average of 26 minutes) in using offline instruction. But EU3 was by far the most successful of the EU's in developing an effective search strategy and in achieving results, as demonstrated by the recall figures. It is therefore clear that we must measure accomplishment, as well as time, to some particular juncture.

Other time differentials of interest are the time between the first search and the first document output -- average six minutes -- and between the first document output and the first useful document found -- average 10 minutes. While there is considerable variation among the individual EU's for these measures, the average figures give us at least some indication of the time and effort required to develop the search in these respects.

An objective measure of accomplishment might be related to how busy the users are in terms of number of commands issued, or number of pages of typescript generated. When these figures are normalized with respect to time, we see that the EU's averaged 0.9 commands per minute, and 15 pages per hour. There is some degree of variation of these figures over the different EU's. EU2 measured lowest in both categories and, in fact, can be said to have accomplished least in both understanding and retrieval results. However, the next-lowest measures are for EU3, who did the best in terms of system understanding, effective use, and retrieval results. The highest figures were attained by EU1; but, a review of her session indicates that she may have missed some opportunities by moving too fast without taking sufficient time for reflection. Therefore, while these measures of busyness may have some correlation with accomplishment, it

is not always necessarily a positive correlation.

Before giving our final conclusions on how good the interface was in helping users learn to use commands, and on what the potential is for improvement, we shall look more carefully at the accomplishments of the EU's in terms of retrieval results and search strategies.

3.4.4 Search Strategy Considerations

The analysis of the retrieval effectiveness in terms of recall for a given effort in time spent, compared EU results with those of analyst searching. This analysis led to a number of tentative conclusions:

1) Users can get some relevant information solely through the interface techniques.

2) The natural-language, keyword/stem approach to searching is one important element in making the interface techniques successful.

3) Although there was a considerable amount of satisfaction by EU's, the retrieval effectiveness in terms of recall percentage was only moderate-to-low.

4) Improvement to interface capabilities is possible through enhancements and extensions to existing techniques.

In order to justify these positions, we shall look at individual sessions.

The session of EU1 proved to be instructive from the viewpoint of search strategy formulation and will be reviewed here in some detail. The topic for EU1 was "the interface between social service agencies and public libraries and how their information sources complement each other". EU1 was considering this topic for a research project, but she had not yet done any searching. She had, however, browsed for about 10 minutes through the ERIC thesaurus of descriptors (ERIC, 1972) before coming to the experimental session, and had brought with her a paper on which she had written the following ERIC descriptors as potentially useful to her search:

1. libraries
2. librarians
3. social services

4. community services
5. human services
6. social agencies
7. public libraries
8. library reference services
9. library services
10. community information services
11. information sources

EU1 started out searching on ERIC (see Table 5 for list of search-related commands) using the keyword stem approach suggested by CONIT with the search on the word stem "librar:". Unfortunately, as described above, the delay message at this point confused EU1 and she did not wait for search completion. Thinking that it was the stemming that caused the delay problem, EU1 reverted to full-word searching (U3 in Table 5). She then tried two-phrase searches (U4 and U5). The first got null results because it is a mixture of two ERIC descriptors, but not a descriptor itself; the second was aborted because of the delay-message problem.

At this point, EU1 went back to the keyword search strategy which she employed for the remainder of the session. EU1 then retrieved and looked at (U12 and U16) the standard document output for documents from the searches "libraries: AND social:" and "public: AND libraries: AND social:". No documents appear very relevant and EU1 decided that she needed to bring in the "interaction" concept. One part of the problem is her mistake in using set numbers which prevented the results from being intersected with the search on "service:".

Following through on her hunch, EU1 searched on the word "interaction:" and intersected the result with the previous searches (U19); this left just two documents. On seeing the title of one of these two, "Performance Guidelines for Planning Community Resource Centers", she felt that she had verified her hunch. She then requested a search on "cooperation" as a synonym for "interaction" and created (U22) SET 18, which is the combination of "cooperation: AND public" AND libraries: AND social:". On looking at the standard output for the first five documents, of the 30 documents in this set, EU1 felt that four of these

Table 5. EU1 Search Strategy

	COMMAND ---->	RESULT
U1.	PICK 56A -->	Connected to ERIC on DIALOG
U2.	FIND LIBRAR -->	DELAYED RESPONSE
U3.	FIND LIBRARIES -->	SET 2 (7447 Documents)
U4.	FIND COMMUNITY SERVICE AGENCIES -->	SET 3 (0 Docs)
U5.	FIND SOCIAL SERVICE AGENCIES --->	DELAYED REPONSE
U6.	FIND PUBLIC --->	DELAYED RESPONSE
U7.	FIND SOCIAL -->	SET 6 (44,534 Docs)
U8.	COMBINE SET 2 AND SET 6 --->	SET 7 (526 Docs)(LIBRARIES: AND SOCIAL:)
U9.	FIND SERVICE -->	SET 8 (37,572 Docs)
U10.	COMBINE SET 6 AND SET 7 -->	(= SET 7 -- mistake, meant SET 6 and SET 8)
U11.	COMBINE SET 2 AND SET 9 --->	SET 10 (= SET 7 -- continuation of above mistake)
U12.	SHOW ---->	Standard information for first five documents of SET 10 shown.
U13.	FIND PUBLIC --->	SET 11 (41,738 Docs)
U14.	COMBINE SET 2 AND SET 11 ---->	SET 12 (2,732 Docs)
U15.	COMBINE SET 10 AND SET 12 --->	SET 13 (222 Docs)(PUBLIC: AND LIBRARIES: AND SOCIAL:)
U16.	COMBINE SET 9 AND SET 12 ---->	SET 14 (= SET 13)
U17.	SHOW ----->	Standard information for first five documents shown.
U18.	FIND INTERACTION --->	SET 15 (9401 Docs)
U19.	COMBINE SET 14 AND SET 15 --->	SET 16 (2 Docs)(INTERACTION: AND PUBLIC: AND LIBRARIES: AND SOCIAL:)
U20.	SHOW --- >	Standard information for the two documents shown
U21.	FIND COOPERATION --	SET 17 (7155 Docs)

Table 5 (continued)

U22. COMBINE SET 14 AND SET 17 ---> SET 18 (30 Docs) (COOPERATION: AND PUBLIC: AND LIBRARIES: AND SOCIAL:)

U 23. SHOW ---> Standard information for first five documents shown

U 24. S S18 ALL D1-30 OFF ----> Offline print for SET 18 (all information) requested

U 25. PICK 55B ---> Connected to LISA on ORBIT

U 26. FIND PUBLIC ----> SET 1 (4379 Docs)

U 27. FIND LIBRARIES ---> SET 2 (11,709 Docs)

U 28. COMBINE SET 1 AND SET 2 ---> SET 3 (3667 Docs)

U 29. FIND COMMUNITY ----> SET 4 (343 Docs)

U 30. FIND SERVICE -----> SET 5 (3837 Docs)

U 31. COMBINE SET 4 AND SET 5 -----> SET 6 (145 Docs)

U 32. FIND COOPERATION --- > SET 7 91178 Docs)

U 33. COMBINE SET 3 AND SET 6 ---> SET 8 (85 Docs)

U 34. COMBINE SET 8 AND SET 7 ---> SET 9 (10 Docs)

U 35. SHOW ----> Standard information for first five documents shown

U 36. S S9 ALL D1-10 OFF ----> Offline print for SET 9 from LISA search requested (all information on all 10 documents)

these documents are either highly or moderately relevant; this judgment is based on seeing titles such as "Cooperation Among Unlike Institutions for Today's Learning Force", and "Community Problems in Five West Central Counties".

Believing from this sample test that she had found a sizeable number of relevant documents, EU1 requested (U24) that all information for all documents be printed offline, and she then proceeded to search another database: LIBINFO (LISA). On this database, EU1 directly and confidently executed a keyword searching strategy; this contrasted strongly with the fumbling development of such a strategy on the first database. One variation is the substitution of "community" for "social" as a search term; we did not determine a reason for this substitution, but the use of this word in the titles of apparently relevant documents seems to be a factor. The set resulting from this strategy, SET 9, has nine documents, and as Table 6 shows, EU1 rated the first five "highly relevant" on the basis of the standard information from (U35). She then requested an offline printout of all the information for all nine documents. Feeling she had retrieved all the documents she needed for this early stage of her research, and not wanting to "use up too much valuable computer time", EU1 then terminated the online session.

When analyzing this session, we looked at the offline printouts that EU1 had requested. In contrast with the standard information requested online, these printouts included abstracts. On the basis of this analysis, we came to two hypotheses:

1) the documents retrieved were not actually as relevant as EU1 thought; and
2) the two keyword stems "librar:" and "communit:" by themselves seemed to be good in a coordinated search strategy.

Searches based on the strategy of (2), with some variations for analysis purposes, were made in ERIC and LISA, plus nine other databases. Full-record output of the resulting documents, or sampled subsets where the retrieved sets were too big, were presented to EU1 for relevance judgment. The searches and judgments are shown in Table 6.

TABLE 6. Search Analysis for EU1

DATABASE	SEARCH	NR	NE	MODE	Relevance H	M	L	N	Precision P_H	P_M	Totals H_T	M_T
ED (ERIC)	A	30	5	UN	2	2	1	0	.4	.8	12	12
	A	30	5	UF	0	0	5	0	.0	0	0	0
	A	30	29	UF	3	9	8	9	.1	.41	3	9
LIBINFO (LISA)	B	10	5	UN	5	0	0	0	1.	1.	10	0
	B	10	5	UF	2	2	1	0	.4	.8	4	4
	B	10	9	UF	2	5	1	1	.22	.78	2	5
TOTAL USER SEARCHES		40	38	UF	5	14	9	10	.13	.8		
ED (ERIC)	C	1612	40	AF	10	7	12	11	.25	.42	403	282
	D	17	17	AF	1	5	8	3	.06	.35	1	5
LIBINFO (LISA)	C	387	40	AF	12	7	12	9	.30	.48	116	68
	E	191	20	AF	9	4	4	3	.45	.65		
	F	91	10	AF	0	1	2	7	.0	.1	0	9
NTIS	C	229	20	AF	6	6	6	2	.3	.6	68	68
SOCSCI	C	73	20	AF	1	4	15	0	.05	.25	4	14
SOC.ABS	C	25	25	AF	0	1	8	16	0	.04	0	1
PAIS	C	14	4	AF	0	1	2	1	0	.25	0	1
	G	17	17	AF	0	1	3	14	0	.06	0	1
INSPEC/EE	C	19	19	AF	1	1	12	5	.05	.1	1	1
SSIE	C	107	20	AF	0	3	6	11	.0	.15	0	16
	H	1079	10	AF	0	0	2	8	0	0	0	0
LIBCON/E	C	172	20	AF	2	2	10	6	.1	.2	17	17
CDI	C	30	20	AF	3	3	5	9	.15	.3	5	5
	H	1225	10	AF	0	0	3	7	0	0	0	0
PSYCH ABS.	C	8	8	AF	0	0	2	6	0	0	0	0
TOTAL ANALYST SEARCHES (EXCL. DUPLICATES)											547	413

LEGEND FOR TABLE 6

SEARCH

A: public: AND libraries: AND social: AND cooperation:

B: public: AND libraries: AND community: AND service: AND cooperation:

C: communit: AND librar:

D: (social: AND (servic: OR agenc:)) OR (human: AND servic:))
AND librar: AND NOT communit:

E: [C] AND NOT servic:

F: [C] AND NOT librar:

G: communit: AND inform:

H: librar:

NR = Number of documents retrieved by search

NE = Number of documents evaluated by EU

MODE = Mode of search and evaluation: U = user search; A = analyst search;
N = online evaluation; F = offline evaluation

RELEVANCE: Number of documents evaluated with high (H), medium (M),
low (L) and no (N) relevance

PRECISION: $P_H = H/NE$; $P_M = (H+M)/NE$

TOTALS: Estimated total number of documents of high (H_T) or medium (M_T)
relevance based on (extrapolation from) those evaluated.

The relevance evaluations and other reaction by EU1 confirmed the two hypotheses. As shown in Table 6, EU1 downrated the relevance evaluations from the ratings made online on the basis of the limited information seen at that time. Apparently, EU1 had simply assumed (or hoped) that the presence of the search words on the retrieved documents would imply the desired relationships among these words. When this assumption was consistent with at least one interpretation of the titles, EU1 took that as confirmation of the assumption.

In fact, however, the "cooperation" mentioned in these documents was generally intra-library cooperation, not cooperation between the libraries and social service agencies. What is more, the words "social", "service", and "community" generally did not refer to agencies other than libraries. Furthermore, there appears to be relatively little, if anything, accessible in the available databases specifically on library cooperation with social-service agencies. The best that can be done for this topic, and what EU1 turned to, is to redefine the topic of interest as "the services offered by libraries which support social service agency efforts (whether or not there is any specific library-agency interaction)".

The analysis of this session supports several of the assertions made in deciding how to assist inexperienced users to search heterogeneous databases. In the first place, controlled-vocabulary searching is generally more difficult and less effective than free-vocabulary searching, especially for the inexperienced user. Even EU1, a library-science student, had trouble trying to use the ERIC thesaurus. As a matter of fact, there are two ERIC thesaurus descriptors that, when intersected, give a moderately good (high-precision, low-recall) search strategy: public libraries and community information services. One problem, especially for the inexperienced users, is the difficulty of finding the good terms and determining that they are good. Analysis shows that intersection of these two search terms, while yielding a high precision (>0.9) causes recall to drop under 10 percent of its value on the more optimum keyword search. There appears to be no way to bring the recall back up to the

keyword search level without using many additional descriptor terms -- many more than the 11 found by EU1. Other terms noted from relevant documents, but that were found by EU1 keyword search A but that would be missed by this simple descriptor search, include: "information centers" "community health services", and "information sources". Of course, as more terms are included in the union (ORing) of the social-services concept search, precision will decline.

The controlled-vocabulary type of search has even more severe problems for this topic, in trying to extend it to most other databases. LISA is one database that does use the descriptors "public libraries" and "community information services"; however, as in ERIC, the intersection search on these two terms seems to yield a very low recall (< 0.1) in LISA of the estimated 293 relevant documents. A brief analysis of other descriptor terms used in LISA uncovered no descriptor search strategy, even a complicated one, that would yield anything close to moderate recall ($\sim .5$) at moderate precision ($\sim .2$).

The NTIS database descriptors are somewhat similar to those of ERIC. However, the reports in NTIS are essentially a subset of those in ERIC for this topic, and so searching this database in any mode does not aid retrieval once ERIC has been searched. The controlled-vocabulary of ERIC carries over poorly, if at all, in any of the other eight databases searched. As with the LISA database, no effective substitute controlled-vocabulary search could be found in any of these databases, but the keyword strategy was reasonably effective for each of them.

An optimum, or even very effective, free-vocabulary keyword search strategy is not necessarily easy to develop. It took a fair amount of analyst time to determine that the search "communit:" and "librar:" is close to optimum. It takes test searches and document output to determine that words like "cooperation", "service" and "social" are more hurtful than helpful when used in addition to, or in place of, the optimum search words. Also, it takes a redetermination by the end user that libraries other than those designated "public" (e.g., municipal,

county, college, etc.) may provide services that are relevant to the topic and, therefore, the limitation to "public" libraries is unnecessarily restrictive. While such analysis may be more or less necessary in different situations for either modes of searching, our analysis of searching for EU1 and the other EU's indicates that the free-vocabulary, keyword/stem approach is easier, especially when searching across muliple databases.

The figures show the need for multiple databases to achieve high recall. Using only one database reduces recall by 0.28 . The top two databases miss eight percent of the recall base. Fully six databases are needed to avoid missing important numbers of documents. Even small numbers can be important if they bring in a different perspective or document type. It may be noted, however, that the precision of the keyword search and the relevance of the relevant documents (ratio of H's to M's) goes down in the peripheral databases.

Using the offline instructional materials, EU1, who did have some prior knowledge of the databases, was able to select most of the significant ones. She was also able to select the two most useful databases in rank order for the online session. Two databases she selected, Psychological Abstracts and Public Affairs Information Service (PAIS), had negligible information on her topic. She failed to select three general databases with moderate amounts of information: LIBCON, SSIE (Smithsonian Science Information Exchange) and CDI (Comprehensive Dissertations Index).

Several searches (see Table 6, searches D, E, F, G and H) were run to determine if the actual recall base was much larger than the one determined by the basic two-word stemmed search. Results of these searches suggest that the actual recall base is probably within 20 percent of that found with all of the analyst searches. Search F shows that a concept subsumed by a database (here "library" in LISA) may be left out of the search statement in which case recall is increased by five percent at the expense of a 0.08 percent drop in precision.

Several additional points on the search strategy can be mentioned. In the first place, truncation searching on word stems clearly aids retrieval; "librar:" picks up "library" and "libraries"; "communit:" retrieves "community" and "communities"; "servic:" covers "service", "services", and "servicing", etc. Secondly, there are possibilities for special search techniques for improving precision; for example, excluding documents on "community colleges libraries" can raise precision significantly in some of the databases without reducing recall noticeably. Increasing the coordination level does increase precision significantly, but at a major cost in recall (witness searches A and B).

The detailed review of the analysis for EU1 given above exemplifies the kind of search strategy analysis accomplished for the other EU's. We shall now give highlights of search strategy considerations for the other EU's.

EU2

In view of the tentative and exploratory nature of the topic as perceived by EU2, we did not perform an extensive search strategy analysis. EU2 hit upon the quite reasonable search term "wings:" for his topic , "wing-section design". In the Engineering Index database, clearly the best one for this topic, this search gives a precision of 0.25, and presumably, a very high recall.

Coordinating the search "design:" with "wings:" raises precision to approximately 0.6 at the cost of halving the recall. Contrarily, it is interesting to note that a truncated search on the stem "wing:" raises recall by about 80 percent, with relatively minor losses in precision. Similarly, truncated searches add significantly to exact-match searching; for example, "design:" recalls 30 percent more than "design".

A controlled-vocabulary searching strategy is possible here using two terms from the Subject Headings in Engineering (SHE,1972) thesaurus: "Wings and Airfoils" and "Design". Searching on these terms instead of the individual words in the basic index appears to raise precision somewhat, at the expense of a serious degradation in recall; for example, the search on "design" as a descriptor recalls only 28 percent as much

as the search on the word "design" in the basic index.

It also appears considerably simpler for a user to select words from the natural language than to have to look up the appropriate headings in a thesaurus, especially (as is the case here) where the thesaurus is not available online. Note that this keyword/basic-index approach also avoids forcing on the user considerations such as the distinctions between main headings and subheadings, controlled terms (descriptors), and free terms (identifiers), descriptor phrases and descriptor words, and title and abstract words. We may also note that these distinctions are further confounded by a few typographical errors and/or spelling variations that have gotten into the database index for the descriptors; some examples are "winga and airfoils", "wings ad airfoils", "wings and aerofoils", and "wings and air foil". While there are only one or two documents involved in each variation, each such variation appears as a separate entry in the online index and, collectively, they tend to clutter up the index and make the selection of search terms more obscure.

EU3

The two databases searched by EU3, SSCI (Social Sciences Citation Index) and SSIE offer an interesting contrast in indexing. SSCI has very shallow indexing (title words only), whereas SSIE has very deep indexing (up to 200 words or more per document), including an elaborate, hierarchical, controlled-vocabulary thesaurus of subject terms.

EU3 carefully read both offline and online instruction on search strategy formulation and moved directly to what turns out to be an effective, simple keyword/stem search for this topic: "world: AND model:". This search gives 112 documents, of which 70 are rated H and 13 M, for a precision of 0.74. Substituting "dynamic:" for "model:", EU3 found 25 additional relevant documents out of 37 retrieved. However, the use of "international:" for "world:" and "simulat:" (e.g., simulation) for "model:" in three other searches found only three more relevant documents. EU3 felt that he had fairly well exhausted the possibilities of effective

subject searching with the above searches. Later analyst searching bore this out. Although citation searching finds an additional estimated 400 relevant documents, there appears to be no simple subject searching strategy that would retrieve many of these with reasonable precision. Citation searching, which was not available to the EU's in the virtual mode, was rather effective for this topic, which has a few key authors such as J. W. Forrester, D. H. Meadows, and C. W. Churchman. Citation searches on these three authors yield 1301 documents with a precision of 0.3. It is estimated that the actual recall base for SSCI might be as much as twice as large as that calculated from the searches done. While it would be quite difficult to retrieve many of these "hidden documents" with subject searching, we suspect that a sizable number could be retrieved with citation searches on other authors.

The search "world: AND model:" in SSIE retrieves 606 documents with low precision (estimated 7%). Coordinating with the word stems "social:" or "econom:" raises precision to the 0.2 to 0.3 range with very little loss in recall. This illustrates the possible need for more highly coordinated search strategies in databases that are deeply indexed, if the simpler strategy retrieves too many irrelevant documents.

EU3 was seeking to achieve as high recall as possible and he wanted to see even those documents he had rated L (low relevance). In this situation, at least 14 databases are seen to be important. Based on the offline instructional material, EU3 had selected seven of the more important ones (SSCI, SSIE, CDI, ENERGY, ENVIRONMENT, LIBCON, and INSPEC-EE) although he had time online to search only the two most important ones. EU3 selected only one database (GRANTS) that later proved relatively fruitless. Analysis uncovered seven other databases of moderate utility (estimated as having more than 10 documents of some relevance): NTIS, PAIS, ERIC, BUSINESS (INFORM), SOCIOLOGICAL ABSTRACTS, MANAGEMENT, and OIL (TULSA). Five other databases were found to have some fewer numbers of relevant documents.

We note that for this topic searching on controlled-vocabulary terms as such, for those databases where they do exist seems generally fruitless.

It is also of interest to note that EU3 was hampered in following CONIT instructions to find appropriate search terms in SSIE because the "PRINT FULL" command into which CONIT regularly translated the "SHOW ALL" request does not output index terms in this one ORBIT database. This problem, now fixed, is another example of how handling individual database peculiarities can be important.

EU4

Due to various CONIT bugs and a system crash (see Appendix D), EU4 retrieved only one relevant document on his topic. It appears that EU4 would have achieved good success with his attempt at following the keyword approach if system problems had not intervened. Keyword-stem searching appears to be a highly efficient technique for avoiding the complexities of the multiple and overlapping controlled- and free-vocabulary indexing used in the prime database for this search: INSPEC-PHYSICS. For example, if EU4 had been able to redo -- as he had started to do -- his search: "helium: AND ion" AND electron excitation:" as "helium: AND ion: AND electron: AND excitation:", he would have retrieved 28 relevant documents instead of one, although precision would have dropped from 0.33 to 0.04.

Precision could be enhanced without much loss of recall, especially at high relevance, by further coordinating with the concept "cross section". What separated the more highly relevant documents from those less relevant were often such intangible or hard-to-search concepts as (1) a more comprehensive analysis, or (2) energy levels (e.g., relativistic effects not wanted).

EU4 eventually was able to follow online instructions with effectiveness, although he never did use SHOW ALL, which could have helped expose indexing usages. Online instruction easily led EU4 to select the PHYSICS database, which was his main source in printed form for regular library searching. He also used the SCIENCE-CITATION INDEX (SCI) database online, because he knew the citation feature could help with "forward (in time) chaining"; however, he never did get to use that feature. Based on his library experience, he thought the CHEMISTRY database was "too spotty" to

be worth searching. However, his later evaluations on analyst searches did indicate the number of moderately relevant documents could be doubled by searching this database, EU4 had some interest in high recall for his thesis bibliography, but he was more interested in trying to insure that he had not missed any very relevant documents.

EU5

EU5 did moderately well in developing and executing a search strategy but he was hampered by three kinds of problems. The first kind involved system bugs and the second kind related to his difficulty in following certain CONIT instructions. Both kinds of problems -- see Appendix D for details -- had the effect of slowing EU5 down somewhat and irritating him, but were not crucial, in themselves, in hindering search strategy development.

The third kind of problem, which had two aspects, was in the area of search-strategy development itself. The first aspect of this problem was his continued attempts at full-phrase searching (e.g., "hearing loss", "abnormal hearing", and "interaural time".) Although none of these was successful ("hearing loss" would have worked in MEDLINE, but was tried only in SCI), and his keyword searching was at least moderately successful, he seemed unconvinced that phrase searching was a poor strategy. Analysis of this session suggests some reasons for this reluctance to stay with keywording:

1) Subject phrases are "natural" to most users, based on their experience with manual systems.
2) The SHOW INDEX response showed that phrases were used (EU5 would have been more successful if he chose only such phrases and not simply invented ones that seemed reasonable.)
3) The CONIT bug in selecting index term tags for searching (see description under EU4 session in Appendix D) cropped up when EU5 tried to select a tag for a one-word term; this failure may have inhibited him from single-word searching, to some extent.
4) EU5 did not avail himself of some explanations (especially, E FIND MORE) that would have emphasized single-word searching.

Undoubtedly, these considerations are not unique to this EU; in fact, they were observed in other experimental sessions.

The second difficulty in search-strategy formulation derives from EU5's failure to use enough synonymous terms in searching or to collect them in "concept bundles" with the appropriate Boolean connectives. Three component concepts can be derived for this topic:

1) hearing
2) the binaural aspect of hearing
3) impairment to hearing .

Terms synonymous or otherwise related to each of these concepts, and later used in analyst searches, are:

1) hearing; hear; sound; binaural; deaf; listen; ear; aural; interaural; dichotic;
2) binaural; dichotic; lateral; locat:, localiz:, localis:, mask:, interaural;
3) impair:, deaf:, defect:, aid:, loss.

EU5 used about half of these terms, but because of failure to search by keywords only, and to group synonyms by the Boolean OR (union operator) before ANDING (intersecting), only a very few of the 200 combination triples were searched. Analysis showed that most of these terms, and many of the combinations, were needed to avoid serious recall and/or precision failures. Some precision problems were noted in using truncated forms (e.g., "heart" matches "hear:" and "earth" matches "ear:") but false drops tend to be excluded on coordination.

EU5 had a bifurcated criterion for determining relevance. On the one hand, he was most interested in those documents where the three concepts were explicitly treated; on the other hand, he also declared relevant those documents with a good treatment of the binaural hearing aspect alone if he felt they could be helpful in analyzing the hearing-impairment aspect. Approximately 10 percent of the estimated 1200 documents in the recall base were in the former, more restricted, category. Even so, these 120 represent a greater-than ten-fold increase in number over the 11 documents EU5 had in a bibliography he had acquired before the computer searching.

Only a few of the 31 relevant documents that EU5 found online fall into the first category. Many others from both categories were included in the searches that EU5 did do, and he could have raised his recall from 0.03 to about 0.1 if he had simply dumped some of these searches for off-line output. That he did not do so may be attributed to several causes.

1) frustration from the various problems encountered online;
2) not wanting to extend the online session unduly;
3) realization that the precision for the category-1 type relevance was rather low; and
4) knowledge that the experimental supervisor/analyst might get this output for him anyway.

EU5 was able to select five highly relevant databases from online information: SCI, MEDLINE, LLBA (Language and Language Behavior Abstracts), Psychological Abstracts, and CDI; he searched the first two of these online. Other databases yielding (more than 40) relevant documents were: Physics Abstracts, SSIE, and NTIS.

In order to achieve high recall in these databases, it is necessary to leave out the impairment concept. (Actually, since LLBA implies a language/hearing concept, only the terms related to the binaural concept were used to achieve high recall in that database). With that strategy, precision ran approximately 0.16 to 0.4. When the third concept is coordinated, precision is raised approximately 10 to 50 percent (actually, by higher percentages if just category 1 relevance is measured), while recall is cut by factors ranging from two to 20 in most of the databases. In SCI, where the indexing is very shallow, recall is cut from 160 to three.

EU6

EU6 presents another instance in which a bifurcated (specific-general) topic became evident. The specific topic is "oral examinations as a testing technique in medical education". The more general topic is "curriculum design philosophy in medical education". The relevance and recall figures in Table 2 reflect a combination of these two topics. The main problem for EU6 was a failure to keyword, due to an over-emphasis on controlled vocabulary. EU6 got two documents when his search "oral examinations" in ERIC

matched that free-vocabulary (identifier) phrase. This may be considered either lucky (in that he got any hits at all on a free phrase) or unlucky (in that a null result might have pushed him into needed keywording). EU6 noted in a post-session interview that he specifically resisted CONIT's suggestions to do keywording because he was "afraid that searches on terms like 'oral' and 'examinations' would be too broad." He went on, instead, to try to use alternate controlled-vocabulary (ERIC thesaurus) terms, with which he was previously familiar. This approach is a dead end for that topic. His search "oral communication AND evaluation methods", for example, yielded documents concerned with the evaluation of oral communication (other than for examination purposes).

A simple keyword/stem expansion of EU6's initial phrase (oral: AND exam:) yields 649 documents at about (a poor) 0.03 precision. Also, the stem "test:" must be added as a synonym for "exam" or about 30% recall is lost. In order to bring up the precision acceptably, one must coordinate with a set of OR'ed terms signifying medical or other higher education. The stem "medic:" is better (about 20 percent higher recall) than the Thesaurus term "medical education"; however, about 40 percent will be lost if one does not OR in an additional series of terms such as "higher education", "colleg:". "professional education", "business education", "physician", etc. It should be noted that EU6 himself got a start on such a precision-enhancing technique. Also, precision is aided by NOTing "language" to avoid documents on oral language exams.

Actually, a much simpler and more efficient strategy would be to insist on an adjacency match between the stems "oral:" and either "exam:" or "test:". This raises precision to over 20% by itself (NOTing "language" would probably raise it to over 50%) without losing more than a few relevant documents. However, the CONIT language does not yet offer that option and, in any case, it is not executable as such in the MEDLINE or ORBIT systems.

The optimized search-strategy formulation in MEDLINE presents an interesting contrast. Starting with the same basic keyword/stem search, one needs for precision to bring in some aspect of education (e.g., search term "educat:" -- medical or higher education need <u>not</u> be specified, since

the database implied medicine) and/or negate dentistry (stem" "dent:") -- the chief irrelevancy generator in MEDLINE in place of "language" in ERIC.

It is possible to achieve some success with a controlled-vocabulary approach to the broader topic on curriculum design, as EU6 attempted, but a keyword approach is still much better. The stem "curricul:" is much simpler and just as effective as OR'ing the approximately 35 ERIC Thesaurus terms containing that stem -- at least several of which are required for good recall. EU6 cut his recall close to zero by insisting on the term "educational philosophy". He wanted documents that discuss the philosophy of curriculum design, but this term is just not widely applied in ERIC indexing.

Acceptable levels of precision are obtained by coordinating "curricul:" with "medical, professional, OR health ed." for ERIC and with just the stem "educat:" for MEDLINE. (Note: "educational philosophy" as an additional coordination term actually reduces precision at the moderate and lower levels of relevance, while reducing recall by a factor of 100 or more.) There may still be more documents in these sets that EU6 wants. Extracting the more relevant ones would involve further precision devices: i.e., title-word searching.

Additional CONIT features that might have helped EU6 are:

a) E FIND MORE
b) SHOW OFFLINE (at least to dump some longer lists for search-strategy review).

The databases selected by EU6 using CONIT explanations are clearly the best ones, in the order selected (ERIC, then MEDLINE).

3.5 Evaluation of Results

3.5.1 General Observations

Because of the limited number of users and limited amount of use of the experimental interface, our conclusions must be considered still tentative, at least with respect to some of the details. Furthermore, our investigations have shown that the superficial presentation of quantitative results may be misleading, especially when comparing results from limited usage. However, through an in-depth analysis of the individual experimental usages, which has been summarized above, we have arrived at certain preliminary conclusions which seem justified by the facts uncovered in our investigations up to the present time.

The central fact is that the virtual-system/translating-computer-interface appears to enable inexperienced end users to extract information they need from multiple heterogeneous databases and systems without recourse to human intermediaries. All six EU's were able to find some useful information within a reasonable period of time; the average time to get to the first relevant document reference was 35 minutes. In order to assess properly the potential utility of the interface concept, and the various techniques employed in its implementation, we shall now analyze the experimental results in greater detail.

The basic design and instructional techniques in the experimental interface proved sufficient to enable the users to learn the mechanics of CONIT use fairly quickly. Each EU used each basic command at least once during the course of her or his online session. While there were a number of mistakes made in using the commands, and several of the more specialized commands and explanations received only limited use, we can say that five of the six EU's appeared to develop a mastery of the basic commands by the end of the session at the terminal. This result tends to support our initial hypothesis, that a modular, structured command-language approach to user control over moderately complicated interactive systems is a viable approach, provided that suitable attention has been given to design for simplicity of use and to adequate online instruction. Getting inexperienced end users online to interact successfully with several existing bibliographic retrieval

systems as we have now demonstrated, is an important achievement that some observers of the information-retrieval scene had begun to doubt was possible.

Besides getting users online and using commands in a proper manner, the interface system must be judged on how effective the searching turns out to be. Our analysis shows that this is a complicated question, and many factors need to be considered before a proper assessment can be made. The prime facts are that estimated fractional recall ranged from 0.01 to 0.23 and absolute recall from one to 171 useful document references for online sessions the duration of which ran from 50 minutes to 124 minutes. As mentioned above, the average time to retrieve the first useful document reference was 35 minutes with a range of from 15 to 41 minutes for the different EU's.

The problem in assessing these facts is that there is no well established standard against which to evaluate them. On the one hand, we may say that enabling inexperienced end users, without human intermediaries, to access any amount of information from the given retrieval systems in the relatively short periods of time experienced represents another major accomplishment. Furthermore, there is likely to be a significant class of potential users of bibliographic retrieval systems who find human information specialists either too inaccessible or too awkward to work with, and who have sufficient need and financial resources to be willing to pay a surcharge if required (see Section 4 for discussion of interface costs), so that the interface approach to access would prove a desirable alternative to access via human intermediaries.

On the other hand, the values for recall appear sufficiently low, and the duration of sessions sufficiently long compared to times that have been reported for information specialist searches [see, e.g., Ross (1979) and Elchesen (1978)] that one might call into question the relative effectiveness of searching via the kinds of interface techniques we have been investigating. In what follows, we argue that this kind of objection is premature and, probably, unwarranted -- at least for the potential of

the interface approach. First, we note that many users do not want or need high recall or, in any case, are satisfied with moderate or low recall as long as they get a few relevant documents to fill out a bibliography or to start or continue a chain of research. The six EU's ran the gamut of recall desires. EU2 and EU5 were looking for high recall; EU2 wanted only a document or two, and the other EU's were scattered in the range between these ends of the spectrum. In terms of satisfying their immediate wants and needs, EU1, EU2 and EU3 did quite well, EU5 and EU6 did fairly well and EU4 did poorly, due to a foreshortened session caused by the MULTICS computer crash.

Although recall may not be the top priority for all users, it is still a very important parameter relating to system performance. The ability to achieve high recall may, at times, be traded off to achieve easier and/or faster response (see, for instance, Marcus, 1978). Therefore we should further consider whether recall is adequate in the interface system.

We should note that the recall estimates shown in Table 2 are, in fact, upper bounds in that the analyst searches, on which they are based, are not fully comprehensive. We have estimated that the actual recall figures may be 20 percent to 50 percent lower, depending on the topic involved. On the other hand, it is not clear how well even an expert information specialist serving as an intermediary would do in terms of recall on these topics. Certainly, the analyst searches were at least moderately complicated, and required a fair amount of analysis, on average, to develop. In any case, additional experiments with expert searchers need to be carried out in order to resolve this question adequately.

3.5.2 The Importance of Search Strategy

One clear conclusion from these experiments -- and certainly one that is not surprising to anyone knowledgeable in information science -- is the importance of search strategy formulation. Formulating an appropriate search strategy is obviously an essential for successful searching in general and it appears to be a particularly critical factor in

the success which inexperienced users can achieve in interacting with existing online databases. It seems clear from these experiments that the adoption and execution of a better search strategy is the main element needed to improve search effectiveness -- as measured by recall as a function of time -- from a range of poor-to-fairly good to one of good-to-excellent.

A companion conclusion -- not nearly so obvious a priori, and still needing additional experimental and analytic investigation -- is that a natural-language-based keyword/stem approach to search strategy formulation appears to be the best approach for achieving moderate-to-high recall, especially for inexperienced users. In our experiments, this approach was consistently superior to one based on selection of controlled-vocabulary (thesaurus) terms except in isolated cases where the latter approach may allow for a quick search achieving higher precision (usually at the expense of lower recall) than a keyword search. In at least some cases, the strict thesaurus approach is simply impossible.

The advantages of the keyword/stem approach are especially significant for the class of user we are particularly trying to satisfy: the inexperienced user who may need access to several databases and systems. For such users, it is especially convenient not to have to struggle with unfamiliar controlled vocabularies. Users can start with a natural-language expression of their topic -- the word "natural" is key -- and apply the same basic strategy across several heterogeneous databases and systems, with assistance from CONIT's common command language.

It should be noted that the success evidenced by the experimental users with this approach depended, in part, on two related facts: (1) CONIT provides certain automated aids to searching; and (2) existing online databases now are, for the most part, "keyword indexed" -- i.e., documents are posted under individual words taken from titles, abstracts (where available), and subject heading phrases, as well as the multi-word phrases themselves. The CONIT automatic search aids include truncation searching and searching under all available subject indexes as the default mode of searching. Keyword indexing permits a natural-language approach to finding documents indexed by a controlled vocabulary

as long as there is a reasonable overlap in words from these two forms of expressions (or just in word stems, where truncation searching is employed).

A large measure of the success of the experimental users may be attributed to the emphasis by CONIT on the natural-langugage keyword/stem approach. The main limitation to even higher search success by the users was the inability of CONIT to encourage the users to adopt this approach more completely in their search-strategy formulation.

3.5.3 Instructional Media and Learning Modes

There does not appear to be a clear winner in the contest between online and offline instruction for database selection and search-strategy formulation. Rather, it appears that both forms of instruction should be included in an optimal system configuration. More important than the medium is the content and quality of the instruction. Users generally prefer online instruction, but a sizable fraction also want offline instruction, as in the form of a reference manual which, many feel, gives a more customary, quicker, and easier mode for referring back to previous instructions than, for example, browsing back through lengthy and bulky computer typescripts. In addition, some users prefer to begin learning how to use a computer system by reading a manual before getting online, where time pressure is more keenly felt.

The experimental results do not confirm either mode as being definitely superior. Individual variations in users, their problems, and external factors, such as system problems, clearly outweigh the (often small) differences in effectiveness measures for the two groups of users.

It was found that the information available online concerning database selection was generally sufficient, at least for selecting the most important databases on a given topic. On the other hand, offline information in a printed format could -- and did, on one occasion -- prove useful, especially for users not at all knowledgeable in the databases, and where the index to databases by professional field or topic identifies a database as covering a topic not obvious from the simple description

of the database -- for example, library and information science being covered by the ERIC (education) database. The expected tradeoffs between offline and online forms apply here; e.g., offline can be less expensive in that computer time is not needed to access the information whereas online can be easier to update, especially for disparate, remote terminal locations.

Users of both forms tended to underplay some of the general databases -- such as LIBCON, NTIS, etc. It was shown that all topics required use of multiple databases to achieve high recall, although there usually were one or two databases that covered a large fraction of the recall base. For some topics, up to 10 or more databases seemed necessary to avoid significant recall gaps. These results underscore the need for access to multiple systems and databases as provided for by CONIT.

In regard to instruction for search-strategy formulation, again we find no clearcut evidence that offline instruction improves retrieval results. We do know that users say that such offline instruction is helpful and they desire to have it available in addition to online instruction. The development of good search strategy appears more related to how assiduously and carefully the user reads and follows the instruction, whether presented offline or online. The critical question, rather than the medium of the instruction, may be how much and what kind of instruction should be presented the user before he attempts his first search or is connected to the first database -- or even before he starts learning system commands.

Generally, it appears that the three EU's who used the offline instruction did better in developing the keyword search approach than did the three who did not have such instruction available. In light of our conclusion that developing the keyword approach is crucial to obtaining better search results, it is likely -- assuming the above appearances are borne out by further evidence -- that additional instruction such as that given to Group A EU's offline would be desirable in the stages before users start submitting search requests. This possibility leads to further questions concerning whether it might not

be better if a user's interaction with the systems should be broken up into several parts: e.g., (1) a practice or learning period -- where, for example, illustrative, or "canned", searches could be tried out by the user; (2) an initial search period -- during which the real problem might be first searched; and (3) a secondary search period in which the initial searching might be revised after reviewing results -- including, perhaps, offline output -- from the initial search.

4. PROSPECTS FOR COMPUTER INTERFACES

As explained below, we conclude that there are excellent prospects for both the immediate and long-range terms for computer interfaces to enhance the cost-effective utilization of interactive retrieval systems while satisfying end users' information requirements. However, as shown in the preceding section, improvements to the experimental interface are desirable in an operational implementation of the virtual-system/translating-computer-interface concept. Possibilities for some of these improvements, as well as other considerations related to interface costs and benefits, are discussed in this section.

4.1 Interface Improvements

4.1.1 Reliability

One obvious area for improvement is in reliability of operation, especially in the context of hardware and software subsystems that are not always reliable themselves. The term "robustness" has been used to characterize systems that meet this criterion of reliability in the face of adversity. In our experiments, the problems to users of unreliable system elements were manifest. The difficulties engendered by such unreliability seem especially pernicious for users in the learning stage; we saw several instances of users giving up on a promising technique or line of approach when their initial attempts met with failure due to the unreliability of some system component. While the supervisor who monitored the experimental sessions was able to overcome the session-terminating aspects of most of the problems, there was a residue of problems that inhibited the effective use of the CONIT system by the EU's.

CONIT already has a number of mechanisms that handle user mechanical errors and difficulties with network and host systems. Considerably more could be done in these respects. The question of availability of systems and rerouting is discussed more fully below in Section 4.3.

4.1.2 Automation and Instructional Assistance

In view of the fact that our analysis has shown that search-strategy formulation is the critical area in which progress could lead to improved retrieval effectiveness, we should comment on prospects for improvement in this area. Two principal approaches to making such improvements are seen at this point: (1) improving instruction to users on how to search effectively; and (2) developing additional techniques for automating search-strategy formulation and execution. These two approaches may be viewed as symbolizing two alternate philosophies of assistance to users of interactive systems. Under the first philosophy, the user is in clear control while the main function of the computer system is to provide instruction and marshal information for the user so that he or she is assisted in making the decision, even down to deep levels of detail.

The second philosophy, on the other hand, aims at keeping control as much as possible in the computer interface by automating decisions as well as execution of tasks; as a way to characterize these two philosophies, we may say that the former emphasizes human intelligence, whereas the latter emphasizes artificial intelligence. In the information-retrieval application, we feel strongly that both approaches need to be used and integrated in a coherent fashion in order to achieve an optimal overall system of user assistance. However, more dramatic improvements will result from new automated techniques.

There are already some automated aids to searching in CONIT: e.g., automatic truncation and basic index searching. A number of procedures that CONIT now suggests to the user could be performed automatically. One example would be to perform automatically the hypothesized optimal stem-Boolean-intersection initial search based on a user-given phrase rather than forcing the user to do those implied operations himself. Another example would be to perform automatically a given search on several different databases without the user having to repeat the request for each database.

Strategies other than the optimal initial search can also be automated. Deciding how to do this effectively may require additional investigations in search-strategy theory, as suggested above. Meadow (1978) has initiated one such investigation. One kind of strategy would require a computer clustering and relevance feedback (see, e.g., Doszkocs 1978). Effective results from the keyword searching in a database with controlled-vocabulary indexing may depend on individual word posting (see Section 3.5). If individual word posting is not implemented, it may be possible to achieve the desired effectiveness by a suitable mechanism for finding index phrases having a given word stem. We have proposed (P3) such a mechanism, which we have termed the Master Index and Thesaurus (MAIT). The MAIT would include terms from all the indexes of the various databases plus posting information. This information would be useful in selecting potentially relevant databases as well as search terms.

A partial implementation of this concept has been developed by Battelle (Colombo and Neihoff, 1977) for "switching (controlled) vocabularies". Other partial implementations aimed at the database selection capability have been accomplished as a research vehicle by Williams and Preece (1977) and as an operational tool by the System Development Corporation (1978). The further development of MAIT-like techniques could have a crucial role in improved algorithms for searching.

In the experimental analysis described in previous sections, we reported various instances in which instruction could be improved. Desired improvements range from simple fixes in instructional dialog to major modifications in instructional formats, media, and learning modes as discussed in Section 3.5.3. The online monitoring performed by the supervisor during the experimental sessions points to another mode of instruction that could be highly effective: the online human consultant. This person could assist several users simultaneously through an online dialog which could be initiated by either a direct user request for help or by the consultant's monitoring of many users and the observation of a situation in which consultation with the user could prove beneficial.

As we have said in Section 3, more important than the format or mode of the instruction is its content. Thus, perhaps the most critical question is finding out what, in fact, makes for good search strategies. We hypothesize that the natural-language-based keyword/stem approach is optimal for most searches as the initial search strategy if one starts with a "reasonably good" set of keywords. This hypothesis has received support from our experimental analysis described here as well as previous analyses (see, e.g., Overhage and Reintjes, 1974). However, assuming that this hypothesis is verified, we are still left with important aspects of the critical question to answer. How do we find the good keywords to start with? How do we identify those exceptional cases in which the proposed strategy is not optimal -- or even, no good at all? In these cases, and in the situations beyond initial searching where search refinement is desired, what then is the optimal strategy?

The CONIT explanatory instructions available to users (see Appendix A) present some partial, tentative answers to these questions in the form of various suggestions on searching in different situations. These suggestions derive from work that we and others (see, e.g., Lancaster, 1973; Marcus, 1971; Jahoda, 1974; Oddy, 1977; Oldroyd, 1977) have done in analyzing searching in online systems. However, much more needs to be done to develop those scattered guidelines into a coherent theory leading to optimal strategies for diverse situations.

We can now see the broad outlines of such a theory as mirrored in the kinds of assistance, both automated and computer-assisted-instructional, that a sophisticated interface would provide.

4.2 Interface Comprehensiveness

The experimental CONIT system was designed with emphasis on performing in a common-language virtual-system mode all of the basic retrieval operations needed by the end user to handle most of his needs. While we

have demonstrated that this level of interfacing does appear to provide a definite utility for a class of users, a question for future research is how many of the specialized functions can, and should, be included in the common mode to accommodate the more advanced needs of users whether they are beginners or experts.

There are three considerations that enter into the determination of how comprehensive to make the common or virtual mode. The first consideration is how difficult and costly is it to implement a particular function in virtual mode. Some functions may be extremely difficult and costly, if not impossible, to perform in a given system. Examples are searching based on specified word separation in ORBIT or MEDLINE and word searching of a nonfull-text indexed field in DIALOG. A common approach can be set up for those systems that do implement these functions; to handle the impossibility of carrying out the given function in the other systems the interface could warn the user and, perhaps, suggest--or automatically execute--substutite similar functions. On the other hand, users can be forced to perform certain functions in the host (non-virtual) language through the pass-through mode.

The second consideration is the difficulty caused to users by these non-virtual situations. The third consideration is, simply, the benefit afforded to the user in implementing a particular function in virtual mode.

Even if a function is not implemented on a given system, the interface can, sometimes, be programmed to handle it. For example, CONIT keeps track of the current search set number and does not require the user to specify it even for output requests to a system (DIALOG) that does impose that restriction. A second, more elaborate, example would be sorting output, say by author name. Whether to implement the operation at the interface to make up for the deficiency at the retrieval system is a question of cost-benefits. It is clearly worth it for the first example (automatically handling current set numbers); the second example (sorting) is more problematical.

4.2.1 Data Structures

One area in which the extension of virtuality raises questions for the direction of future interface development is the variability in bibliographic data structure among different databases. In previous reports (P3, P6) we described how this problem impacts both searching and output and the possibility of handling these problems through the mechanism of a common bibliographic data structure into which and from which translations could be performed by the interface, analagous to the translation for the common-command language.

In our research we found it possible to circumvent this problem to a large extent by limiting the search and output options available to the user in the virtual mode; in effect, CONIT presents a very simple common bibliographic data structure and an acceptance of the lack of perfect translation in some cases. For example, the ORBIT standard (default) PRINT output may not always have the same elements as the DIALOG mode 3 output--which we have taken as most closely approximating "standard" output.

While such approximations and limitations have not prevented the experimental interface from achieving a significant level of utility, the question of the value of the further extension of the common bibliographic data structure concept remains to be explored. Such explorations are needed to determine the feasibility of two major enhancements to interface capabilities: (1) merging results from searches on different databases, and (2) recreating database subsets at the interface for direct search there, rather than at a remote host.

4.2.2 Databases and Systems

Another facet of the problem of differences among databases which we have analyzed with respect to our experiments is the variaton in search strategy required to achieve satisfactory results in the different databases. This is related to the formulation of optimized search strategies as discussed above.

Another dimension to the question of interface comprehensiveness is the number of different retrieval systems to be accessible through the interface network. The current CONIT system connects to three of the most widely used bibliographic retrieval systems. There are many other such systems currently available which could add significant retrieval capability to an interface-facilitated network.

Our rule-based approach to interface structure makes the addition of new systems to the network relatively easy: basically, an additional set of translation rules must be added for each new system--the old rules remain unchanged. In cases where the new systems share a close relationship to one already handled--e.g., NASA and ERDA RECON are very similar to DIALOG--relatively few additional rules may be needed. (We note that ORBIT and the two MEDLINE systems already share many of the same rules.) In cases where the new system brings in different functional capabilities, or handles a function in a very different way, relatively more new rules may be needed. The question in adding new systems is not so much of possibility--that seems answerable in the affirmative--but of practicality--what are the cost-benefits tradeoffs.

Beyond purely bibliographic retrieval there is the question of interfacing to databases of all kinds and to more general database management systems (DBMS's) and, in fact, to information systems of various kinds. Some investigations have been begun in this direction (see, e.g., Glaseman and Epstein, 1978; Erickson, et al, 1976; Sagalowicz, 1977; Kameny, 1978). One reason that we have been able to demonstrate the utility of the interface approach as effectively as we have is that we have restricted ourselves to the bibliographic retrieval application. It remains to be seen how increasing generality of function affects the feasibility and practicality of achieving effective virtual-system type networking of heterogeneous interactive systems through the translating interface approach.

4.3 Interface Configurations

We implemented our experimental interface in a large time-sharing computer because this was an effective context for development, modification, and testing. The best configuration for an operational interface system needs to be determined. A greater efficiency could probably be achieved by using a dedicated computer, assuming that such a computer can be kept reasonably busy. A number of investigators (e.g., Anderson, 1976; Goldstein, 1977; and Rosenthal, 1975) have experimented with minicomputer configurations for interface systems.

Some of these investigators foresee the possibility of implementing interfaces on microprocessors and incorporating them directly in each terminal; such a configuration has been called an intelligent terminal. The functional capabilities that are proposed for intelligent terminal interfaces do not, generally, come close to the range and depth of capabilities we have proposed for the virtual-system interface. Our CONIT 3 system required on the order of 200K bytes of storage in MULTICS and an interface with a number of the more advanced capabilities described above would demand more computational resources. Therefore, it is unlikely that a full virtual-system interface would be implemented on an intelligent terminal although some of the capabilities desired in such an interface could well be incorporated in an intelligent terminal.

Because a virtual-system interface involves many scattered users and a number of scattered retrieval systems, the overall configuration involves a network and the network ramifications should be considered. In the first place, it is highly desirable that the interface and the disparate retrieval systems be embedded in host computers in a modern digital communications network so as to be able to take advantage of the speed, efficiency, and reliability such a configuration implies, in contrast with the much lesser capabilities inherent in the ad hoc configuration involving the autocall unit we had installed in MULTICS. MULTICS is, in fact, a host computer on both the TELENET and TYMNET computer networks on which many of the major retrieval systems, including the three accessible by CONIT,

are hosts; however, the software and hardware facilities required to make full computer-to-computer (as opposed to terminal-to-computer) connections have not yet been installed, and so we have not yet been able to take advantage of the improved communications capabilities implied by such packet-switched networks.

In an operational environment, it would be desirable to have a number of interfaces distributed around the network in order to provide greater reliability and efficiency in terms of load sharing and reduced communications requirements. The possible design and configuration of such a distributed interface system has been discussed in one of our previous reports (P6).

4.4 Costs and Benefits

The direct benefits of an operational interface of the kind we have tested have been substantiated by the experiments: simpler access to and use of existing heterogeneous bibliographic retrieval systems by all types of users, especially end users. Such systems could become accessible to a large number of end users who do not now use them because of the awkwardness of engaging expert intermediaries to help them perform a search. This, of course, may lead to reduced costs through economies of scale, as mentioned in Section 1.

Although precise cost estimates cannot be made without a detailed operational design, some order-of-magnitude estimates can be made. The interface required duplication of certain functions regularly performed by retrieval systems: the parsing of input requests and the handling of dialog. Also, communications requirements are roughly doubled, in that the interface-to-retrieval-system links have to be added to the terminal-to-computer links. (It should be noted that these extra communications links, as we have implemented them, generally do not add significantly to the perceived response time for the user.) Some functions -- such as selection of, and command translations into, target systems -- would be unique to the interface. On the other hand, the major component function of the actual storage and retrieval from large databases would not be required within an interface, at

least within a relatively modest-sized interface of the degree of complexity found in our experiments. The CONIT 3 interface system required about 200K bytes of storage in MULTICS.

Summing up, we estimate an additional cost for the computer-system components of approximately twenty percent for this type of interface over those same costs for direct access. The figure of 20 percent increased computer costs is partially supported by observations on costs of the experimental CONIT of from $5.00 to $20.00 per hour. An operational system would be designed more with efficiency in mind, and could be expected to have lower costs than an experimental interface. Another benefit of a self-instructing interface is a reduction in costs for helping new users; such help is a major cost in providing service on retrieval systems.

A more sophisticated interface with many of the advanced features described in this section would be correspondingly more costly, perhaps as costly as one of the retrieval systems itself. However, the benefits would include much improved retrieval capability for all classes of users; it would be expected that a given retrieval effectiveness -- in terms of a certain recall level in a specified time -- would be achieved at reduced cost through advanced interfaces.

It may be noted that certain apparent cost reductions might be achievable with interface techniques beyond what has already been proposed here. One example would be the multiplexing of two or more users on a single line to a host retrieval system; this would cut costs from the retrieval system by a factor of two or more at the expense of some additional complexity in the interface in keeping the searches of two or more users identified uniquely by user, and some increased response time when results are requested simultaneously. Another example would be disconnecting retrieval systems when several minutes of user inactivity in making a request are expected; resumed connection and searching would be done only as requested by the user, and past results would be automatically recreated quickly (without the time-consuming user typing times) as needed.

We have <u>not</u> included such techniques in our projection of cost savings --

although variations on them may present possibilities for savings in user's time -- because they represent savings that are more apparent than real. The cost "savings" here depend on the host system charging by the connect-hour, rather than total computer resource usage. Eventually, we expect retrieval systems to charge more on the basis of total resources used (e.g., including CPU time, amount of online communications, etc.) rather than simply by connect time. When this kind of charging is put into effect, as is now done in many computer systems, resource-intensive search operations, like truncation searching, will cost more -- which is consistent with their greater effectiveness. This may make some interface functions of the kind we have suggested more expensive in an intermediate future time. In the longer-range future, we can expect systems to be redesigned so that such desired functions will be done more efficiently -- for example, by stemming on input and implicit storage of longer sets of document references rather than explicit storage. Improved, and possibly more standardized, retrieval systems may reduce the complexity needed in interfaces; in fact, new systems should probably be designed with interface capabilities built in.

Without regard to the ultimate benefits of operational interface systems, there is a major utility to experimental interface systems as research tools. In this capcity it is possible, as we have demonstrated, to evaluate potential modification to, or enhancements of, retrieval-system features without requiring expensive and disruptive modifications to existing retrieval systems. Thus, for example, proposed command-language standards or search-strategy aids can be evaluated before great expense is involved in implementing what may be less-than-optimal techniques.

5. CONCLUSIONS

In this report, we have described investigations into the concept of a translating-computer-interface/virtual-system mode for assisting users in their access to, and use of, heterogeneous interactive bibliographic retrieval systems. An experimental interface system, based on this concept, has been built and tested under controlled conditions with six end users who had not previously used computer retrieval systems. A detailed analysis of the experimental usages has shown that the users were able to master the basic commands of the interface sufficiently well to find useful document references on topics for which thay had a current information need.

The success of the experimental interface is attributable to a design emphasizing simplicity of use and to a comprehensive collection of instructional aids. An important component of the instruction involved procedures for developing search strategies based on a natural-language keyword/stem approach to searching. It is concluded that operational interfaces employing the limited set of techniques implemented in this research can provide for increased usability of existing retrieval systems in a cost-effective manner. Such interfaces should be especially useful for inexperienced end users and others who cannot easily avail themselves of expert searcher intermediaries. Furthermore, it is concluded that improved search effectiveness for all classes of users is feasible with more advanced interfaces that would include additional techniques which have been suggested by this research but not yet fully designed or tested in experimental interfaces.

6. PROJECT BIBLIOGRAPHY

P1. Therrien, Charles W. Data Communications for an Experimental Information-Retrieval Network Interface . M.I.T. Electronic Systems Laboratory Technical Memorandum ESL-TM-515; August 1, 1973. (NTIS:PB237075).*

P2. Marcus, R.S. "A Translating Computer Interface for a Network of Heterogeneous Interactive Information-Retrieval Systems". Proceedings of the ACM Interface Meeting on Programming Languages and Information Retrieval (November 1973). Association of Computing Machinery, SIGIR Forum. 9(3):2-12; Winter, 1974.

P3. Reintjes, J.F., Marcus, R.S. Research in the Coupling of Interactive Information Systems . M.I.T. Electronic Systems Laboratory Report ESL-R-556; June 30, 1974. (NTIS: PB237974; ERIC:ED94758).*

P4. Marcus, Richard S. "Network Access for the Information Retrieval Application". Panel on Access to Computer Networks, 1975 IEEE Intercon Conference Record, Session 25/4: 1-7, April, 1975.

P5. Marcus, R.S. "Networking Information Retrieval Systems Using Computer Interfaces". Proceedings of the 38th Annual American Society for Information Science Conference. 12:77-78; October 26-30, 1975.

P6. Marcus, Richard S., Reintjes, J. Francis. The Networking of Interactive Bibliographic Retrieval Systems . M.I.T. Electronic Systems Laboratory Report ESL-R-656; March, 1976. (NTIS PB252407; ERIC:ED125533).*

P7. Reintjes, J. Francis. "The Virtual-System Concept of Networking Bibliographic Information Systems". NATO Advisory Group for Aerospace Research and Development (AGARD) Conference Proceedings No. 207 on Advancements in Retrieval Technology as Related to Information Systems. Institute for Defense Analyses, Arlington, Virginia, 20-21 October, 1976. Pp. 9-1 to 9-7.

P8. Marcus, Richard S.; Reintjes, J. Francis. Computer Interfaces for User Access to Heterogeneous Information-Retrieval Systems . M.I.T. Electronic Systems Laboratory Report ESL-R-739: April, 1977. (NTIS:PB269126 ERIC: ED142180).*

P9. Marcus, Richard S.; Reintjes, J. Francis. Experiments and Analysis on a Computer Interface to an Information-Retrieval Network . M.I.T. Laboratory for Information and Decision Systems Report LIDS-R-900; April, 1979 (This report).

P10. Marcus, Richard S.; Reintjes, J. Francis. "A Translating Computer Interface for End User Access to Heterogeneous Information-Retrieval Systems". (In preparation).

* These are numbers for ordering through NTIS and ERIC.

7. REFERENCES

Anderson, Robert H.;Gillogly, James J. "The RAND intelligent terminal agent (RITA) as a network aid", Proc. of the 1976 National Computer Conference, AFIPS Press, pp. 501-509.

Benenfeld, Alan R.; Pensyl, Mary E.; Marcus, Richard S.; Reintjes, J.F. NASIC at M.I.T.: Final Report. M.I.T. Electronic Systems Laboratory Report ESL-FR-587, February 28, 1975.

Caruso, Dorothy Elaine. Training Modules for Use of Scientific and Technical Information Services. Final Report on NSF Grant No. DSI 76-09538, University of Pittsburgh Graduate School of Library and Information Sciences, December, 1977.

Colombo, D.S.;Niehoff, R.T. Improved Access to Scientific and Technical Information through Automated Vocabulary Switching -- An Experiment Based upon Selected Energy-Related Terminology. Battelle Columbus Laboratories, Final Report on NSF Grant No. SIS75-12924, May, 1977.

Curtis, Dade T. "Has the End User Been Forgotten? An Information Profile of Active Research Bioscientists", in The Value of Information (Proceedings of the 6th Mid-year Meeting, May 19-21, 1977, Syracuse, N.Y.) Amer. Soc. for Information Science, pp. 41-45.

Doszkocs, Tamas E. "An Associative Interactive Dictionary (AID) for Online Bibliographic Searching". Proc. of the ASIS 41st Annual Meeting, V15: 105-109, November, 1978.

ERIC (Educational Resources Information Center). Thesaurus of ERIC Descriptors. CCM Information Corporation. New York; 1972.

Erickson, L.R.; Soleglad, M.E.; Westermark, S. L. ADAPT I Preliminary Functional and System Design Specification. Logicon, Inc. Report 76-C-0899-2 November, 1976.

Glaseman, S.; Epstein, H. Design of a Terminal Subsystem Providing User Access to Distributed Data Base Management Systems. RAND Corporation Draft Report WN-10177-ARPA. April, 1978.

Goldstein, Charles M.; Ford, William H. "The User-Cordial Interface". Online Review 2(3): 269-275; September, 1978.

Hendrix, Gary G.; Sacerdoti, Earl D.; Sagalowicz, Daniel.; Slocum, Jonathan. "Developing a Natural Language Interface to Complex Data". ACM Trans. on Database Systems 3(2): 105-147; June, 1978.

Jahoda, Gerald. "Reference Question Analysis and Search Strategy Development by Man and Machine", J. Amer. Soc. for Information Sciences, 25(3): 139-144, May, 1974.

Kameny, I.; et al. EUFID: The End User Friendly Interface to Data Management Systems. System Development Corporation Draft Report. February, 1978.

Kennedy, T.C.S. "Some Behavioral Factors Affecting the Training of Naive Users of Interactive Computer Systems", Internat. J. of Man-Machine Studies, 7(6): 816-834, November, 1975.

Lancaster, F.W.; Fayen, E.G. Information Retrieval On-Line, Melville Publishing Co., Los Angeles, CA., 1973.

Mann, William C. "Why things are so bad for the computer-naive user", Proc. of the National Computer Conference, AFIPS Press, 1975, 785-787.

Marcus, Richard S. "The Importance of Recall". J. of the American Soc. for Information Science 29(6): 314-315, November, 1978.

Marcus, R.S.; Benenfeld, A.R.; Kugel, P. "The User Interface for the Intrex Retrieval System". in Interactive Bibliographic Search: The User/Computer Interface, Donald E. Walker, editor, AFIPS Press, 1971, Montvale, N.J. (pp. 159-201).

Meadow, Charles T.; Toliver, David E.; Edelmann, Janet V. "A Technique for Machine Assistance to Online Searchers". Proc. of the ASIS Annual Meeting. 15:222-225; November, 1978.

Negus, A.E. EURONET Guidelines: Standard Commands for Retrieval Systems. Commission of the European Communities Report DG XIII; December, 1977.

Oddy, R.N . "Information Retrieval through Man-Machine Dialogues", The Journal of Documentation, 33(1): 1-14; March, 1977.

Oldroyd, Betty K.; Citroen, Charles L. "Study of strategies used in on-line searching", On-Line Review 1(4): 295-310; December, 1977.

Overhage, C.F.J.; Reintjes, J.F. "Project Intrex: A General Review". Information Storage and Retrieval, 10(5): 157-188 (1974).

Rosenthal, Robert. "Accessing Online Network Resources with a Network Access Machine", IEEE Intercon Conference Record, Session 25/3, April 1975.

Ross, Johanna C. "Searching the Chemical Literature via Three On-Line Vendors: A Comparison". J. of the Amer. Soc. for Information Science. 30(2): 103-106; March, 1979.

Sagalowicz, Daniel. "IDA: an Intelligent Data Access Program". Proc. of the Third International Conference on Very Large Data Bases. Tokyo, Japan; October, 1977.

SHE (Subject Headings for Engineering). Engineering Index, Inc., New York, N.Y.; 1972.

System Development Corporation. SDC Search Service ORBIT News Newsletter, 6(8): 1-12. October, 1978.

Wanger, Judith; Cuadra, Carlos A.; Fishburn, Mary. Impact of On-line Retrieval Services: A Survey of Users, 1974-75, Systems Development Corporation Report, 1976.

Williams, Martha E.; Preece, Scott E. "Data Base Selector for Network Use", Proc. of the ASIS 40th Annual Meeting. 14:(34); September, 1977.

APPENDIX A

EXPLANATORY MESSAGES REQUESTABLE BY USER

This appendix contains all the online instruction for CONIT 3 which is requestable online by the user through the HELP and EXPLAIN commands. (Other online instruction is provided automatically by CONIT throughout the user's session; see Sections 2 and 3 and Appendix B for details.)

The format of the explanations is that of the offline printed CONIT reference manual. The reference manual was not used by the experimental user in the experiments described in this report; they relied on the online HELP or EXPLAIN commands, or for Group A, the offline instruction shown in Appendix C.

The first explanation lists all explanations available in CONIT 3 and the EXPLAIN command that evokes each explanation. The explanation number is not printed online. Thus, for example, EXPLANATION 5 is evoked by the command "EXPLAIN CONIT" or "E CONIT" and begins as follows:

```
EXPLANATION OF CONIT
CONIT is an experimental ...
```

BEST COPY AVAILABLE

CONIT REFERENCE MANUAL

CONIT EXPLANATION 1

**1

LIST OF EXPLANATIONS

A full listing of all the CONIT explanations that you can request online is given below. In order to see one of these explanations online, type the the appropriate EXPLAIN command, as shown below. These explanations are also printed in order on the following pages of this reference manual.

EXPLAIN COMMAND	EXPLANATION GIVEN
1. E EXPLANATIONS	This list of explanations
2. E START	Starting or restarting your CONIT session
3. E EXPLAIN	How to use the EXPLAIN command
4. E CONCEPTS	Some basic CONIT concepts
5. E CONIT	Short background explanation of CONIT project
6. E CONVERSE	How to converse with CONIT
7. E ERRORS	How to correct typing errors
8. E BREAK	How to use the BREAK key to interrupt CONIT
9. E COMMANDS	List of basic CONIT commands
10. SHOW DATA	List of 7 areas with data bases available through CONIT
11-17. S DATA 1...7	List of data bases in each of the 7 areas
18. E E DATA	How to get information on a particular data base
19. E E FIELDS	How to get detailed field information on a data base
20. E PICK	How to pick a data base
21. E SYSTEMS	Retrieval systems having data bases
22. E PICK SYSTEMS	How to pick systems
23. E SHOW STATUS	How to know what you are currently connected to
24. E DISCONNECT	How to disconnect the current data base and system
25. E STOP	How to stop your CONIT session
26. E FIND	How to find (search for) documents
27. E FIND AUTHOR	Finding documents by author name
28. E SHOW INDEX	Getting valid index terms to search under
29. E FIND INDEX	Searching on terms found by SHOW INDEX
30. E FIND MORE	Broadening searches so as to get more documents
31. E FIND BETTER	Narrowing searches to get documents of higher relevance
32. E COMBINE	Combining sets of retrieved documents
33. E SHOW	Showing information on documents, data bases, etc.
34. E SHOW DOCS	Details on showing document information
35. E SHOW REVIEW	Reviewing current searches
36. E SHOW NEWS	Getting news from connected system
37. E SEND	Sending commands in non-CONIT language
38. E COMMENT	Making comments to CONIT

****

Electronic Systems Laboratory
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139

December 21, 1977

CONIT EXPLANATIONS 2-4

**2
STARTING OR RESTARTING YOUR CONIT SESSION
If you are already connected to CONIT, you may start, or restart, your CONIT session by typing:
start
followed by a carriage return (RETURN) key.

The message you will get after typing START is shown below:

++++CONIT:
Welcome to CONIT. If you make a typing mistake, strike the BREAK key, wait for the user cue (****USER:), and then retype the line.
For more help on how to use CONIT, type
help
followed by a carriage return; otherwise, you may now type any CONIT command.

1800.5 [Time of day]
*****USER:

**3
THE EXPLAIN COMMAND
The name of the basic help command is explain . In using it you may type either the complete word or its abbreviated form e . The command name must always be followed by the item you want explained and by a carriage return. For example, if you type
explain commands
or just
e commands
CONIT will list all the basic commands you can use and a short explanation of them. The inexpert user should now type: e commands

**4
EXPLANATION OF CONCEPTS
To have a concept explained type explain followed by a concept name. Some concepts you can have explained are:

CONCEPT NAME	CONCEPT
commands	List of CONIT commands
converse	Mechanics of the user's conversational dialog with CONIT
co .t	Short background of CONIT system
systems	Info on retrieval systems CONIT can communicate with
data	List of data bases you can search

For each of the above explanations, CONIT will suggest other, more detailed explanations that you can get. For a list of all available online explanations that you can request, type: e explanations
Note that CONIT will automatically (without your asking) give explanations and suggestions during the course of your session if it detects errors on your part or as it thinks that suggestions may be helpful.

****
This page last updated: 77-12-17

CONIT EXPLANATIONS 5-0

**5

EXPLANATION OF CONIT

CONIT is an experimental computer interface that connects you to different information retrieval systems and allows you to select data bases of these systems to search (find references to documents). Search requests for all systems and data bases can be made in one language, the CONIT language.

CONIT will assist you as your searching progresses.

Since the CONIT language does not yet handle all functions of every system, you may prefer to use the language of the connected retrieval system, if you know it, for specialized functions; for info on this point type: e send

CONIT has been developed under research grant from the National Science Foundation Division of Science Information to study networking and other improvements for interactive information systems.

**0

EXPLANATION OF HOW TO CONVERSE WITH CONIT

You converse with CONIT by giving it commands. Each command consists of a command name which may be followed by one or more additional words to make the meaning of the command clear.

To signal the computer that you have completed your command you MUST strike the carriage return key; the computer will not respond until you do.

CONIT will respond to your command with a message. To signal that its message is complete and that it is again waiting for your command CONIT will print the "user cue": *****USER:!

You cannot give a command until you get the USER:! cue but you can interrupt CONIT in its processing of its last command by typing the BREAK key. CONIT will then give you a USER:! cue.

For most errors you make CONIT will give you an informative error message. If you detect an error BEFORE you strike the carriage return key, you can cancel what you have typed so far by striking the BREAK key. After receiving the USER cue you may retype the command correctly.

(If you are connected to MULTICS via ARPANET, send BREAK by typing: as i p)

For information on other ways to edit typing errors, type: e edit

****

This page last updated: 77-12-17

CONIT EXPLANATIONS 7-8

**7

EXPLANATION OF CORRECTING ERRORS

For most errors you make CONIT will give you an informative error message. If you detect an error BEFORE you strike the carriage return key, you can cancel what you have typed so far by striking the BREAK key. After receiveing the USER cue you may retype the command correctly.
(If you are connected to MULTICS via ARPANET, send BREAK by typing: @s i @)

The BREAK key (sometimes labeled ATTENTION (ATTN)) can also be used to interrupt CONIT actions AFTER you request them; for info type: e interrupt

There are two ways to cancel errors without having to BREAK and wait:
1- Type the AT SIGN (@) to cancel the command line up to that point and immediately type the correct line. Thus, if you type

shoe d@show data

you will correctly get the SHOW DATA command.
2- n NUMBER SIGNS (#) will cancel the last n characters you typed. Thus, either of the two lines

shoe#w data
shoe da####w data

will correctly enter the SHOW DATA command. Combinations also work; e.g.,

shoe d###w daat@shop dat@show dato#a

If you are connected to MULTICS thru ARPANET, you must type 2 AT SIGNS (@@)

**8

EXPLANATION OF INTERRUPTING

The BREAK key (also called ATTENTION (ATTN) can be used to interrupt actions and messages before or after you request them by simply depressing the BREAK key ONCE and waiting for CONIT to give you the USER:: cue before making a new request. This will take a variable amount of time depending on how long it takes CONIT to get the other system to "stop talking".

This feature may be very useful to avoid waits where, e.g., you have already seen some instructions, info, or you have seen enough document reference output, or you are simply tired of waiting for an action to start or be completed.
(NOTE: BREAKING occasionally causes systems to drop out or hang up.)
(If you are connected to MULTICS thru ARPANET, send BREAK by typing: @s i @

****

This page last updated: 77-12-17

CONIT EXPLANATION 4

```
**4
     EXPLANATION OF COMMANDS
   The basic commands you will be giving are these (either the
full name or abbreviation may be typed):
 NAME    ABBREV     SHORT EXPLANATION
explain      e      Used to get explanations about using CONIT
pick         p      Used to pick a data base (and retrieval system)
find         f      Used to find documents in a data base
show         s      Used to show information on documents, data bases, etc.
combine      c      Used to combine sets of retrieved documents
send                Used to make a comment or send command in non-CONIT language
stop                Used to stop your CONIT session
   In addition to the regular commands listed above, we note below a few
special operations by CONIT or the user which are important to the dialog:
 NAME          OPERATION/MESSAGE
user cue     The string "*****USER:" is CONIT's cue that the user may type
RETURN       At the end of each command the user must type the RETURN key
               to signal the computer to do the command
BREAK        Strike the BREAK KEY (do NOT type the word "break") to cancel a
               typed line or to interrupt CONIT. Wait for the "****USER:" cue
               before typing a new line
errors       For a detailed explanation of how to correct typing errors
               (with or without using the BREAK key), type: e errors
   More detailed explanations of the above commands and operations can be
had by typing   explain   or   e   , followed by the name of the item you
wish to have explained and a carriage return.  Other items you can have
explained are listed under   e concepts   .
   The inexpert user should now type:
    e pick

**************************
          This page last updated: 77-12-17
```

CONIT EXPLANATIONS 10-11

```
**10
    DATA BASE SUBJECT AREAS
   Listed below are 7 subject areas in which data bases are available
  1.  Physical Sciences (Physics, Chemistry)
  2.  Engineering (including Chemical Engineering and patents)
  3.  Geosciences, Energy, Environment
  4.  Biomedical, Agriculture, Food
  5.  Social Sciences, Politics, Education, Humanities
  6.  Business, Marketing, Management, Grants
  7.  General (all, or most, subject areas covered)
For a list of the data bases in one of these areas, type: show data X
where X is the number of the area: e.g.,
     show data 2
will list the Engineering data bases you can search.


**11
     DATA BASES IN PHYSICAL SCIENCES
   For each data base there is listed:
     CONIT number
     CONIT name
     code letters of systems having the data base
       (D=DIALOG, N=NLM/MEDLINE, O=ORBIT, S=SUNY/MEDLINE)
     short explanation of data base
11  SCI (D) Science Citation Index (SCISEARCH)
12  PHYS (D) Science Abstracts-A: Physics (INSPEC)
13a CHEM77 (D) Chemical Abstracts Condensates 1977- (+CHEMIN indexing)
13b CHEM72 (D,O) Chemical Abstracts Condensates 1972-7b (1972- for ORBIT)
13c CHEM70 (D,O) Chemical Abstracts Condensates (CHEMCON) 1970-71
13i CHEMIN (D) Chemical Abstracts Subject Index Alert (CASIA) (73-7b)
13t CHEMTERM (D) Chemical Name Dictionary: CHEMNAME; see also MEDCHEMTERM, etc
   You may now pick a data base by typing   p X   where X is a number or name.
However, if you need more information on a data base before picking, either
refer to the printed descriptions, if you have them, or type: e data X Y
where X is the data base number or name and Y is a system name. E.g.,
     e data 13b orbit
or
     e data chem72 dialog
For a list of data bases in another of the seven areas, use the
SHOW DATA command again.

**********
          This page last updated: 77-12-17
```

CONIT EXPLANATION 1c

```
**12
     DATA BASES IN ENGINEERING
   For each data base there is listed:
     CONIT number
     CONIT name
     code letters of systems having the data base
       [D=DIALOG, N=NLM/MEDLINE, O=ORBIT, S=SUNY/MEDLINE]
     short explanation of data base
21   ENGIN [D,O] Engineering Index (COMPENDEX, EI)
22   EE [D] Science Abstracts-B,C: Electrical,Electronic,Computer,Control
23   ME [D] Information Service in Mechanical Engineering (ISMEC)
24a  METALS [D] Metals Abstracts/Alloys Index (METADEX)
24b  ALUM [D] World Aluminum Abstracts (WAA)
25a  PATDER [D] Derwent: Central/World Patents Indexes
25b  PATIFI [D] IFI/Plenum: General, electrical, and mechanical U.S. Patents
25t  PATTERM [D] U.S. Patent Office Classification Codes
26a  PATCHEM77 [D] IFI/Plenum: Chemical Patents- U.S. and foreign equiv (1977-)
26b  PATCHEM50 [D] Same as PATCHEM77 for 1950-76
26c  PATCHEMCA [D] Chemical Abstracts Patent Concordance, US and foreign
26p  PAPER [D] Paper Chemistry
   You may now pick a data base by typing   p X   where X is a number or name.
However, if you need more information on a data base before picking, either
refer to the printed descriptions, if you have them, or type: e data X Y
where X is the data base number or name and Y is a system name. E.g.,
     e data 21 orbit
or
     e data engin dialog
For a list of data bases in another of the seven areas, use the
SHOW DATA command again.

*****************************
         This page last updated: 77-12-17
```

CONIT EXPLANATION 13

```
**13
     DATA BASES IN GEOSCIENCES, ENERGY, ENVIRONMENT
   For each data base there is listed:
     CONIT number
     CONIT name
     code letters of systems having the data base
       [D=DIALOG, N=NLM/MEDLINE, O=ORBIT, S=SUNY/MEDLINE]
     short explanation of data base
11   SCI [D] Science Citation Index (SCISEARCH)
23   ME [D] Information Service in Mechanical Engineering (ISMEC)
32   METEOR [D] Meteorological and Geophysical Abstracts (MGA)
33   OCEAN [D] Oceanic Abstracts (NOAA)
33b AQUA [D] Aquatic Sciences and Fisheries Abstracts (ASFA)
34   ENVIR [D] Environmental Abstracts (ENVIROLINE)
35a POLLUT [D,O] Pollution Abstracts
35b AIRPOLLUT [D] Air Pollution (APTIC)
36   ENERGY [D,O] energy Information and Environment Abstracts (ENERGYLINE)
37a OIL [D] Oil and gas exploration, development and production (TULSA)
37b OILNEWS [D] Oil and energy news (P/E NEWS--Am. Petroleum Institute)
   You may now pick a data base by typing   p X   where X is a number or name.
However, if you need more information on a data base before picking, either
refer to the printed descriptions, if you have them, or type: e data X Y
where X is the data base number or name and Y is a system name. E.g.,
     e data 35a orbit
or
     e data pollut dialog
For a list of data bases in another of the seven areas, use the
SHOW DATA command again.

*******************************
          This page last updated: 77-12-17
```

CONIT EXPLANATION 14

```
**14
     BIOMEDICAL, AGRICULTURAL, AND FOOD DATA BASES
   For each data base there is listed:
     CONIT number
     CONIT name
     code letters of systems having the data base
       (D=DIALOG, N=NLM/MEDLINE, O=ORBIT, S=SUNY/MEDLINE)
     short explanation of data base
11   SCI (D) Science Citation Index (SCISEARCH)
41a  AGRIUS (D,O) U.S. National Agricultural Library (AGRICOLA)
41b  AGRICOM (D) Commonwealth Agricultural Bureaus (CAB)
41c  FOOD (D,J) Food Science and Technology Abstracts (FSTA)
41r  AGRIRES (D) Agricultural Research (USDA-CRIS)
42   BIOSIS (D,O) Biological Abs. and Bioresearch Index 1972- (ORBIT: 1974-)
42b  BIO69 (D) BIOSIS 1969-73
42t  BIOTERM (D) BIOSIS codes and synonyms dictionary
43   MEDLINE (S,N) Index Medicus, by National Library of Medicine (NLM)
43m  MEDMON (N) NLM-cataloged monographs (CATLINE)
43n  MEDNOW (S,N) Current month of MEDLINE
43s  MEDCHEMTERM (S,N) Chemical and medical terms (CHEMLINE)
43t  MEDTERM (S,N) Medical Subjects Headings Dictionary (MESH)
43v  MEDAV (N) Medical Audio-Visual materials (AVLINE)
44   TOX (N) Toxicology (TOXLINE)
44d  TOXEFF (N) Toxicity effects data (NIOSH-RTECS)
44t  TOXTERM (D) Toxicological terms (TOSCA)
45   EPIL (N) EPILEPSYLINE
46   CANCER (N) Cancer Therapy and Carcinogenesis Abstracts (CANCERLIT)
47r  CANCERRES (N) Cancer research projects (CANCERPROJ)
   You may now pick a data base by typing   p X   where X is a number or name.
However, if you need more information on a data base before picking, either
refer to the printed descriptions, if you have them, or type: e data X Y
where X is the data base number or name and Y is a system name. E.g.,
     e data 41a orbit
or
     e data agrius dialog
For a list of data bases in another of the seven areas, use the
SHOW DATA command again.

**********
         This page last updated: 77-12-17
```

CONIT EXPLANATION 15

```
**15
     DATA BASES IN SOCIAL SCIENCES, POLITICS, EDUCATION, HUMANITIES
   For each data base there is listed:
     CONIT number
     CONIT name
     code letters of systems having the data base
       (D=DIALOG, M=NLM/MEDLINE, O=ORBIT, S=SUNY/MEDLINE)
     short explanation of data base
51   SOCSCI (D) Social Sciences Citation Index (SOCSCISEARCH)
52   SOC (D) Sociological Abstracts
52b  CHILDAB (D) Child abuse and neglect
53   PSYCH (D) Psychological Abstracts
54a  FED (D) FEDERAL INDEX to government activities
54b  CRECORD (D) Congressional Record
54c  PUBLIC (D) Public Affairs Information Service (PAIS)
54n  FEDNEW (D) Most recent month of FED records BEFORE loading into FED
55a  LANG (D) Language and Language Behavior Abstracts (LLBA)
55b  LIBINFO (D) Library and Information Abstracts (LISA)
56a  ED (D,O) Educational Resources Information Center (ERIC)
56b  EDEXCHILD (D) Exceptional Child Education Abstracts
56c  EDAV (D) Curriculum planning media materials (NICEM)
56d  VOTECH (D) Vocational and technical education
57a  HIST (D) Historical Abstracts (non U.S.)
57b  AMHIST (D) America History and Life (AHL), U.S. and Canada
58   ART (D) Modern art and design
   You may now pick a data base by typing   p X   where X is a number or name.
However, if you need more information on a data base before picking, either
refer to the printed descriptions, if you have them, or type: e data X Y
where X is the data base number or name and Y is a system name. E.g.,
     e data 55a orbit
or
     e data ed dialog
For a list of data bases in another of the seven areas, use the
SHOW DATA command again.

**********
          This page last updated: 77-12-17
```

CONIT EXPLANATION 16

```
**16
     DATA BASES IN BUSINESS, MARKETING, MANAGEMENT, GRANTS
   For each data base there is listed:
     CONIT number
     CONIT name
     code letters of systems having the data base
       (D=DIALOG, N=NLM/MEDLINE, O=ORBIT, S=SUNY/MEDLINE)
     short explanation of data base
61  BUS (D,O) Business, management, marketing, etc (ABI/INFORM)
61b BUSCAN (D) Canadian Business Periodical Index
62a MANAGE (D) Management, finance, industrial relations, etc
62b ACCOUNTANT (D) Accounting, auditing, investments, taxation, etc
63a MARKETAB (D) PREDICASTS: Worldwide company and product info
63d MARKETDOD (D) Government contracts, RFP's, etc in defense areas
63i MARKETIN (D) PREDICASTS: Indexes to MARKETAB
63n MARKETNOW (D) Current month records BEFORE loading into MARKETAB
64a STATPLANT (D) EIS: Info on 117,000 U.S. business establishments
64b STATUS (D) PREDICASTS: U.S. industry and government statistics
64c STATUSTIME (D) PREDICASTS: Annual Time Series for U.S. industry
64d STATREG (D) PREDICASTS: Time Series for U.S. metropolitan areas
64f STATFRN (D) Similar to STATUS for foreign statistics
64g STATFRNTIME (D) Similar to SATUSTIME for foreign statistics
65a GRANTS (D) Grant programs by governmental, commercial, and private groups
65b FONDIR (D) FOUNDATION DIRECTORY (non-governmental, nonprofit)
65c FONGRANTS (D) Grants given by foundations in FONDIR
66a CHEMNEWS (D,O) News from Chemical Industry Notes by Chem Abstracts
66b DRUGNEWS (D,O) Pharmaceutical News Index
   You may now pick a data base by typing   p X   where X is a number or name.
However, if you need more information on a data base before picking, either
refer to the printed descriptions, if you have them, or type: e data X Y
where X is the data base number or name and Y is a system name. E.g.,
     e data 61 orbit
or
     e data bus dialog
For a list of data bases in another of the seven areas, use the
SHOW DATA command again.

**********
          This page last updated: 77-12-17
```

CONIT EXPLANATIONS 17-19

**17

DATA BASES COVERING ALL, OR MOST, SUBJECT AREAS

For each data base there is listed:

CONIT number
CONIT name
code letters of systems having the data base
[D=DIALOG, N=NLM/MEDLINE, O=ORBIT, S=SUNY/MEDLINE]
short explanation of data base

71 NTIS [D,O] Government-sponsored R&D reports (Natl Tech Info Service)
72e LIBCONE [O] Library of Congress books and monographs in English
72f LIBCONF [O] Library of Congress foreign (non-English) materials
73 SSIE [O] Smithsonian Science Information Exchange (current research)
74 THESES [O] Comprehensive Dissertation Abstracts (doctoral theses)
75 NEWSCAN [O] Canadian News Index

You may now pick a data base by typing p X where X is a number or name. However, if you need more information on a data base before picking, either refer to the printed descriptions, if you have them, or type e data X Y where X is the data base number or name and Y is a system name. E.g.,

e data 71 orbit

or

e data ntis dialog

For a list of data bases in another of the seven areas, use the SHOW DATA command again.

**18

EXPLAINING DATA BASES

In order to get an online explanation of a data base, type: e data X where X is the CONIT number or name of the data base.

A given data base may exist on two or more systems, in which case the implementations and explanations will be different. If you want to get the explanation for a particular system, add the system name to your command. E.g.,

e data ed dialog

and

e data ed orbit

will get explanations of the ED data base from DIALOG and ORBIT, respectively. If you do not name a system, CONIT will pick one for you, as in the PICK command.

If you are not connected to a system from which you want an explanation, CONIT will connect you (after disconnecting from the current system and data base, if any). The data base you are currently connected to will not be changed when you ask for an explanation, unless the explanation you request is from a system other than the one you are currently connected to. In that case you will be connected to the default data base of the explaining system.

**19

EXPLANATION OF FIELDS

Fields are categories of information in a data base (e.g., title, author, document source, report number, index terms, etc.) that can be used for searching or for showing documents found as the result of searching. For simple searching (FINDing) and SHOWing in the CONIT language you do not need to know field names of the data bases, but for specialized searching or showing, which must be done in the language of the retrieval system, you must use the data base field names. To see these use the E FIELDS command after the E DATA command.

****

This page last updated: 77-12-21

CONIT EXPLANATION 20

**20

EXPLANATION OF PICK COMMAND

The PICK command is used to pick a data base to search in. To pick a data base, type pick X or simply p X where X is the number or name of the data base; for example:

pick 43

or

p medline

will get you connected to the MEDLINE data base. For an online list of the data bases you may pick, type:

show data

To switch from one data base to another, type pick Y where Y is the number or name of the new data base. Note, however, that switching generally results in previous results being dropped by the computer.

In order to save on connect-time charges you may, at times, want to disconnect the current data base; to do this, type:

disconnect

or its abbreviation

dis

Again, disconnecting generally loses previous results. After disconnecting you must PICK again to resume searching.

CONIT will automatically connect to a system that has the data base you picked. For an explanation of how you can override CONIT's choice of system, type: e pick system

If the above explanations are on paper, we suggest that the inexpert user now pull up the paper, tear off the instructions for reference, and then pick a data base (after getting information on the data bases, if needed).

****

This page last updated: 77-12-17

CONIT EXPLANATIONS 21-23

**21
EXPLANATION OF SYSTEMS
CONIT can connect to 3 different retrieval systems: 1) MEDLINE of the National Library of Medicine; 2) Lockheed DIALOG (LRS); and 3) System Development Corporation (SDC) ORBIT.
MEDLINE, whose main data base is also named MEDLINE, specializes in medicine. It also has other data bases in related biomedical fields. Both DIALOG and ORBIT have many data bases covering all disciplines.
For a list of data bases available in the different systems, type: s data
If one system is not available, or to make more complete searches, you may want to try several systems and data bases.
There are two versions of the MEDLINE system: NLM and SUNY. The NLM version has a few data bases not available at SUNY. For more info pick a third system: NLM NEWS. When you 'pick medline' CONIT tries to get SUNY. You may get NLM by typing: pick medline nlm
CONIT uses either the TELENET or TYMNET network to get a system. If a system is not available thru one network, you may try the other; e.g.,
pick medline telenet
or
pick medline nlm telenet

**22
EXPLANATION OF PICKING SYSTEMS
To connect to a data base CONIT must first connect to a retrieval system that has the data base. Some data bases are available on more than one system. CONIT will automatically select a system for you unless you prefer to override CONIT's choice. To do this, specify both data base AND system; for example,
pick ntis orbit
connects you to the NTIS data base in ORBIT, while
p 71 dialog
connects you to the NTIS data base in the DIALOG system.
If you specify a system but no data base, you will be connected to the default data base for that system. For example, the default data base for DIALOG is ED.
For more information about the systems and networks, type: e systems

**23
SHOWING STATUS
The SHOW STATUS command tells what data base and system you are currently connected to, if any.

****
This page last updated: 77-12-17

CONIT EXPLANATIONS 24-25

**24

EXPLANATION OF DISCONNECTING

The DISCONNECT command (abbreviation 'dis') can be used to disconnect the currently connected retrieval system and data base. This will leave you connected to CONIT but will avoid any additional charges from the retrieval systems. Disconnecting will drop from the computer any current search sets (offline requests will NOT be affected, however).

To disconnect from CONIT (as well as any current retrieval system), use the STOP command.

**25

EXPLANATION OF STOP COMMAND

The command 'stop' will log off CONIT after disconnecting any system that you may currently be connected to.

Before you stop CONIT we would appreciate any comments you may have. For explanation type: e comment.

You may disconnect a retrieval system WITHOUT stopping CONIT, type: e dis

****

This page last updated: 77-12-17

CONIT EXPLANATION 26-27

**26

EXPLANATION OF FIND COMMAND

The FIND command is used to search for documents indexed under a particular term. Type find X where X is the term you are searching for. In general, use terms that are specific to your topic. Two examples:

```
find erythrocyte
find radiation
```

We recommend that you start by searching on single (key) words or word stems and then combine the resultant sets as needed. E.g., the 3 commands:

```
find computer
find memor
combine set1 and set2
```

will find documents about COMPUTER MEMORIES by getting documents indexed under BOTH COMPUTER or COMPUTERS, etc., (set1) AND under MEMOR(Y,IES) (set2). For additional explanations that may be useful for searching type:

```
e combine        (for combining sets of retrieved documents)
e find author    (for searching by author name)
e find more      (for ways to find additional documents)
e show index     (for browsing through the valid index terms)
e find index     (for searching on terms given by SHOW INDEX)
e find better    (for ways of getting more relevant documents)
```

**27

EXPLANATION OF AUTHOR SEARCHING

In addition to general subject searching, you can request a search specifically on a particular author. To do this type: find author followed by the author's name in the form: [last-name, initial]. E.g.,

```
find author martin, r
```

Which can be abbreviated

```
f au martin, r
```

will do a search on authors with the last name 'Martin' and a first name beginning with 'R'.

If this gives too many 'false drops' (e.g., you get documents by Richard Martin when you wanted only those by Robert Evan Martin), you can be more specific as in the following two examples:

```
f au martin, robert
f au martin, r. e
```

Note, however, that different data bases vary as to how they format author names. Some use only initials; some use periods after initials, others do not.

The best strategy may be to start with a broad author search and SHOW some documents. If you then want to restrict the search, use the author name format that you see in the SHOW output.

****

This page last updated: 77-12-17

CONIT EXPLANATIONS 28-30

**28

SHOWING INDEX TERMS

The 'show index X' command lists terms alphabetically near the term X in the index for the data base you are currently connected to. The number of documents posted under each term is also given. Thus,

show index avogadro

will show if AVOGADRO is a proper search term in your current data base (you can check spelling). You will also see how many, if any, documents are posted under that term, or alternately, other related phrases like Avogadro's number or rule or constant, etc.

For some data bases you will also be given an indication of whether there are listings for related thesaurus terms.

**29

FINDING WITH TERMS FROM SHOW INDEX COMMAND

Terms that are shown by the SHOW INDEX command may, of course, be used as search terms in the FIND command.

If you are connected to DIALOG, you can select particular terms from the displayed list by designating their code tags instead of the full spelled-out form of the terms. Thus,

find e7

will do a search on term 7 in the SHOW INDEX display, while

find e1-e8,e12

will do a search on the first 8 terms plus term 12 (results of the 9 searches are combined with OR, as for synonyms).

**30

FINDING MORE DOCUMENTS

Some ways to find additional documents:

1. Use 'show index X' command to find additional search terms
2. Use 'or' in COMBINE command to collect documents under synonymous terms
3. Get synonymous or related terms from 'show all' document output, or from thesauri, dictionaries, indexes, your head, etc.
4. Break a search phrase into single words
5. Use other data bases and systems
6. Use truncated (stemmed) search terms. CONIT searches on all terms that BEGIN with your search term. Thus, 'find computer' will retrieve documents indexed under 'computer', 'computers', 'computer programs', 'computerized', etc. To broaden your search, reduce the length of your search term. E.g., 'find comput' will get all of the documents from the 'computer' search PLUS those indexed under 'compute', 'computation', etc.
7. Drop less important terms from search: e.g. 'effects' in 'heat effects'

****

This page last updated: 77-12-17

COMIT EXPLANATIONS 31-32

**31

FINDING BETTER DOCUMENTS

If you have too many non-relevant documents, your search terms may not be specific enough or you may not have included an important concept. To correct this, search on the concept you have omitted; then combine the new and previous search sets using COMBINE with AND.

If a major part of your search set is not relevant and you can isolate the part by some search, you can exclude the non-relevant part by combining the original result AND NOT the new search. E.g.,

find memor
find comput
combine set1 and not set2

will remove (at least some of) the computer memory documents. However, it may also remove some documents on human memory where computers are used.

If you know the retrieval system language, you can also do special searches as by date, report number, "important terms" only, etc. E.g., in DIALOG:

find rn=nasa-cr-129524
find amorphous(W)state/ti

Some ORBIT or MEDLINE examples:

find new (sa)
find *lead and *poisoning

**32

EXPLANATION OF COMBINE COMMAND

The COMBINE command allows you to make combinations of the sets of documents you have previously found from searching your currently connected data base; for example,

combine set 2 and set 5

will yield a new set that contains only documents which are common to both sets 2 AND 5. ANDing is useful in narrowing the search to your specific topic. Alternately, if want to broaden your search, use OR; e.g.,

combine set 2 or set 5

merges the 2 sets into a single set keeping all documents from either set. ORing is useful in combining results from synonymous terms. The 3 commands:

find sulfate
find sulfite
combine set 1 or set 2

will collect documents about sulfates OR sulfites into one set. Also,

combine set 2 and not set 5

will make a new set which contains documents in set 2 but not in set 5.

****

This page last updated: 78-9-5

CONIT EXPLANATIONS 33-34

**33

EXPLANATION OF SHOW COMMAND

The show command gives information about documents that have been found in searching, about data bases, about index terms to search on, etc.

To have CONIT show standard citation information on some of the last set of documents you have found type 'show'. You may also be more specific; e.g.,

show set 3 title documents 1-4

will cause the titles of the first 4 documents of set 3 to be shown to you.

show all d 2

gives ALL information available on the SECOND document of the current set.

To have the SHOW output printed offline and mailed, add 'off'; e.g.,

s s3 all d1-57 off

will get all info for the first 57 documents of set 3 printed offline. I'd' is the abbreviation for 'document(s)', 's' for 'show' and 'set'). For more details on how to get document information type: e show d Examples of other information that can be obtained are given below:

COMMAND	BRIEF EXPLANATION
show data	Lists data bases currently available through CONIT
show index X	Lists terms near X in index for current data base
show review	Reviews search sets created so far on current data base
show status	Tells what data base and system are connected
show news	Gives news from connected system

For more details on each, type: e show Y

where Y is the command you want explained; e.g.,

e show index

**34

SHOWING DOCUMENT INFORMATION

The basic command for showing information on documents found has the form:

show documents 5-7 set 3 title

which causes the titles of documents 5, 6, and 7 of set 3 to be output. (You may abbreviate: 's' for 'show' or 'set' and 'd' for 'document(s)'.)

If you do not specify the SET parameter, you will get the last set you have found.

If you do not specify the DOCUMENT (d) parameter, you will get the first few documents in the set (usually about 5, if there are that many).

You can get all document information by substituting 'all' for 'title'. If you specify neither 'title' nor 'all', you will get the standard citation information which includes title, author, and locator information. Different systems and data bases give you somewhat different types of information for the categories TITLE, ALL, and CITATION. This is especially true of other-than-bibliographic data bases like FOUNDATION GRANTS or CONGRESSIONAL RECORD.

To save money and online time you may want to have your SHOW output printed offline (NOT at terminal) and mailed. To do this add the word 'off'; e.g.,

show off all s3 d1-57

means print offline all info on the first 57 documents in set 3.

****

This page last updated: 77-12-17

CONIT EXPLANATIONS 35-38

**35

SHOWING SETS IN REVIEW

The command 'show review' lists all the sets you have found in searching the currently connected data base with a short description of each set, and, for some systems, the number of documents in each set.

**36

SHOWING NEWS

The command 'show news' will give you the news from the system you are currently connected to. This is in addition to the short broadcast news you may have gotten when you first connected to the system.

News for the MEDLINE systems is only available by picking a special nlm news system by typing: pick nlm news

**37

EXPLANATION OF THE SEND COMMAND

To send a command to a system in its own language you may use the SEND command. E.g.,

send .file 12

will connect you to DIALOG file 12 WITHOUT dropping sets just found.

send t2/1/1-20

will give just format 1 output (doc numbers) for DIALOG set 2

send "print indented ti, an

will give title and doc number info for MEDLINE and ORBIT.

We welcome your comments on CONIT. To make a comment use the 'send comment' command. For explanation type: e comment

**38

COMMENTING

We would like to get your comments on the good and bad features you find in using CONIT to aid us in our study of information systems. To make a comment use the "comment" command for each line of comment: e.g.,

comment the common language feature is good but show is a bad name

comment print or type is a better name than show

****

This page last updated: 77-12-17

APPENDIX B

EXAMPLE CONIT SESSION

As explained in Section 2.2, we present in the Appendix excerpts from the online session for an experimental user (EU3) as an example of typical interactions that occur between users and the CONIT system.

During the session, the user issued 90 commands. These are labeled U1, U2, ..., U90. The CONIT responses are labeled C1, C2, ..., C90. Two of the commands, U5 and U6, were not actually issued by EU3 but have been added for completeness as explained in Section 2.2. Additional explanatory information on this session is found in Section 2.3.4 and Appendix D.

The dialog has been retyped in a format close to that which actually appeared on the original typescript. For a comparison, see Fig. 3 in Section 2; Fig. 3 is a photoreproduced excerpt of the typescript from command U19 to U26. In this Appendix, material left out is indicated by ellipses (...) or by a note indicating where the full CONIT explanatory message can be found in Appendix A.

U1: START

C1: ++++CONIT:
WELCOME TO CONIT. IF YOU MAKE A TYPING MISTAKE, STRIKE THE BREAK KEY,
WAIT FOR THE USER CUE (****USER::), AND THEN RETYPE THE LINE.
FOR MORE HELP ON HOW TO USE CONIT, TYPE
HELP
FOLLOWED BY A CARRIAGE RETURN; OTHERWISE, YOU MAY NOW TYPE ANY CONIT
COMMAND.

0956.6
*****USER::

U2: HELP

C2: ++++CONIT:
THE NAME OF THE BASIC HELP COMMAND IS EXPLAIN . IN USING IT YOU
MAY TYPE EITHER THE COMPLETE WORD OR ITS ABBREVIATED FORM E . THE
COMMAND NAME MUST ALWAYS BE FOLLOWED BY THE ITEM YOU WANT EXPLAINED
AND BY A CARRIAGE RETURN. FOR EXAMPLE, IF YOU TYPE
EXPLAIN COMMANDS
OR JUST
E COMMANDS
CONIT WILL LIST ALL THE BASIC COMMANDS YOU CAN USE AND A SHORT
EXPLANATION OF THEM. THE INEXPERT USER SHOULD NOW TYPE: E COMMANDS

0.957.6
*****USER::
U3: E COMMANDS

C3: [See EXPLANATION 9 from Appendix A]

U4: E PICK

C4: [See EXPLANATION 20 from Appendix A]

U5: SHOW DATA

C5: [See EXPLANATION 10 from Appendix A]

U6: SHOW DATA 5

C6: [See EXPLANATION 15 from Appendix A]

U7: PICK 51

C7: ++++CONIT:
IN ORDER TO CONNECT TO THE DATABASE YOU PICKED, SOCSCI (NUMBER 51),
CONIT WILL CONNECT TO THE DIALOG SYSTEM.
TO MAKE THIS CONNECTION CONIT WILL FIRST PHONE THE TYMNET NETWORK.
THE WHOLE CONNECTION PROCESS WILL TAKE ABOUT A MINUTE OR TWO.
PLEASE WAIT FOR THE USER CUE (*****USER::) BEFORE MAKING A NEW COMMAND.
NOTE THAT THE BREAK KEY CAN BE USED AT ANY TIME TO INTERRUPT CONIT
AS WELL AS TO CANCEL A LINE: WHEN YOU STRIKE THE BREAK KEY CONIT WILL
ABORT WHAT IT'S DOING; YOU MAY GIVE A NEW COMMAND AFTER YOU GET
THE USER CUE.

C7:(cont'd)

PHONE CONNECTION MADE TO TYMNET.
TYMNET RESPONDING.
LOGGING ONTO DIALOG.
YOU ARE NOW CONNECTED TO THE LOCKHEED DIALOG RETRIEVAL SYSTEM--
YOU ARE NOW CONNECTED TO THE SOCSCI (NUMBER 51) DATA BASE.
FOR EXPLANATION OF FINDING DOCUMENTS TYPE: E FIND

1006.1
*****USER::

U8: E FIND

C8: [See EXPLANATION 26 from Appendix A]

U9: E COMBINE

C9: [See EXPLANATION 32 from Appendix A]

U10: E FIND MORE

C10: [See EXPLANATION 30 from Appendix A]

U11: E FIND INDEX

C11: [See EXPLANATION 29 from Appendix A]

U12: E SHOW INDEX

C12: [See EXPLANATION 28 from Appendix A]

U13: E FIND BETTER

C13: [See EXPLANATION 31 from Appendix A]

U14: E FIND AUTHOR

C14: [See EXPLANATION 27 from Appendix A]

C15: PHONE LINE DETACHED
THE PHONE CONNECTION AND RETRIEVAL SYSTEM HAVE BEEN DISCONNECTED
FOR OTHER SYSTEMS THAT MAY HAVE THE INFORMATION YOU WANT, TYPE: E SYSTEMS

```
U16:  PICK 51
      . . .

U17:  SHOW INDEX MODEL

C17:  ++++CONIT:
      REF     INDEX-TERM          TYPE ITEMS  RT
      E1     MODALITY---------------    149
      E2     MODALS-----------------      5
      E3     MODDERMAN--------------      1
      E4     MODE-------------------    267
      E5     MODECATE---------------      1
      E6    -MODEL------------------   6355
      E7     MOLELE-----------------      1
      E8     MODELED----------------     19
      E9     MODELER----------------      1
      E10    MODELES----------------      2
      E11    MODELING---------------    636
      E12    MODELIZATION-----------      4
      E13    MODELL-----------------      7
      E14    MODELLANSTALT----------      1
      E15    MODELLE----------------      1
      E16    MODELLER---------------      2
      E17    MODELLING--------------     82
      E18    MODELLKINDERGARTEN-----      1
      E19    MODELO-----------------      1
      E20    MODELR-----------------      1
      TO SEE MORE TERMS TYPE:  SHOW MORE
      TO SELECT ANY OF THESE TERMS FOR SEARCHING TYPE: E FIND INDEX

      1026.0
      *****USER::

U18:  SHOW INDEX WORLD

C18:  ++++CONIT:
      REF    INDEX-TERM          TYPE ITEMS RT
      E1     WORKWEEK---------------    14
      E2     WORKWEEKS--------------     2
      E3     WORKWIFE---------------     1
      E4     WORKYARDS--------------     1
      E5     WORLABY----------------     1
      E6    -WORLD------------------  6520
      E7     WORLDLINESS------------     1
      E8     WORLDLY----------------     6

             . . .
```

U19: FIND MODEL

C19: ++++CONIT:
NO RESPONSE RECEIVED FROM LAST TRANSMISSION.
SHOULD CONIT WAIT FOR A RESPONSE ANY LONGER? (ANSWER YES OR NO)
RETRIEVAL SYSTEMS MAY BE SLOW IN RESPONDING AT TIMES. YOU SHOULD ANSWER YES TO THE ABOVE QUESTION ABOUT 3 SUCCESSIVE TIMES (UNTIL ABOUT A MINUTE HAS ELAPSED) AFTER WHICH TIME YOU MAY ASSUME EITHER THE RETRIEVAL SYSTEM OR CONIT OR THE NETWORK CONNECTION BETWEEN THE TWO IS NOT WORKING PROPERLY. AT THAT POINT YOU SHOULD ANSWER NO TO THE QUESTION. THEN TRY YOUR REQUEST AGAIN. IF YOU STILL GET NO RESPONSE, DISCONNECT FROM THE RETRIEVAL SYSTEM (TYPE DIS) AND TRY TO PICK IT AGAIN (POSSIBLY USING A DIFFERENT NETWORK) OR PICK A DIFFERENT SYSTEM.
*****USER::

U20: YES

C20: NO RESPONSE RECEIVED FROM LAST TRANSMISSION.
SHOULD CONIT WAIT FOR A RESPONSE ANY LONGER? (ANSWER YES OR NO)
*****USER::

U21: YES

C21: FROM HOST:
YOUR SEARCH RESULTS IN SET 1 WHICH CONTAINS 10415 DOCUMENTS.
TO SEE REFERENCES TO THE FIRST FEW DOCUMENTS TYPE: SHOW

1034.5
*****USER::

U22: FIND WORLD

C22: ++++CONIT:
NO RESPONSE RECEIVED FROM LAST TRANSMISSION.
SHOULD CONIT WAIT FOR A RESPONSE ANY LONGER? (ANSWER YES OR NO)
*****USER::

U23: YES

C23: NO RESPONSE RECEIVED FROM LAST TRANSMISSION.
SHOULD CONIT WAIT FOR A RESPONSE ANY LONGER? (ANSWER YES OR NO)
*****USER::

U24: YES

C24: FROM HOST:
YOUR SEARCH RESULTED IN SET 2 WHICH CONTAINS 6904 DOCUMENTS.
TO SEE REFERENCES TO THE FIRST FEW DOCUMENTS TYPE: SHOW
FOR AN EXPLANATION OF HOW TO COMBINE SETS TYPE: E COMBINE

1036.7
*****USER::

U25: COMBINE SET 1 AND SET 2

C25: ++++CONIT:
YOUR SEARCH RESULTED IN SET 3 WHICH CONTAINS 113 DOCUMENTS.
TO SEE REFERENCES TO THE FIRST FEW DOCUMENTS TYPE: SHOW
TO REVIEW YOUR SEARCH SETS TYPE: SHOW REVIEW

1037.9
*****USER::

U26: SHOW

C26: ++++CONIT:

3/3/1
785089
HOUSING REQUIREMENTS AND NATIONAL RESOURCES - IMPLICATIONS OF UN WORLD MODEL (EN)
LAKSHMANAN TR; CHATTERJEE L; ROY P
JOHNS HOPKINS UNIV, CTR METROPOLITAN PLANNING & RES/BALTIMORE/MD/21218; JOHNS HOPKINS UNIV, DEPT. GEOG & ENVIRONM ENGN/BALTIMORE//MD/21218; GALLAUDET COLL/WASHINGTON//DC/20002
HABITAT, U2, N3-4, P277-289, 1977

3/3/2
783666
ECONOMIC-MODELS OF WORLD AGRICULTURAL COMMODITY MARKETS - COCOA, COFFEE, TEA, WOOL, COTTON, SUGAR, WHEAT, RICE - ADAMS, FG AND BEHRMAN, JR (EN)
JONES GT
JOURNAL OF AGRICULTURAL ECONOMICS, U28, N2, P326-326, 1977

3/3/3
777057
BEHAVIORAL-MODEL FOR HEALTH PROMOTION AND DISEASE PREVENTION (EN)
NG LKY; MANDERSCHEID RW; DAVID DL
NIMH/WASHINGTON//DC/20032; WORLD MAN FUND/WASHINGTON//DC/
JOURNAL OF ALTERED STATES OF CONSCIOUSNESS, U3, N1, P95-95, 1977

3/3/4
772792
BARILOCHE MODEL - 3RD WORLD DISASTER STUDY PREDICTS (EN)
JAHNKE M
ENVIRONMENTAL POLICY AND LAW, U3, N2, P74-75, 1977

3/3/5
759229
HUMANIZATION OF LABOR WORLD - NECESSARY CONDITIONS, CONCEPTS AND MODELS - GERMAN - PREISS, C (GE)
TRAUTWEINKALMS G
GEWERKSCHAFTLICHE MONATSHEFTE, U28, N7, P469-469, 1977

```
      YOU MAY NOW SEE OTHER INFORMATION ON YOUR SEARCH RESULTS; TYPE: E SHOW
      FOR EXPLANATION OF OTHER ACTIONS YOU MAY NOW DO TYPE:  E COMMANDS

      1044.1
      *****USER::

U29:  SHOW SET 3 ALL D1-113 OFF

C29:  ++++CONIT:

      PRINTED 3/5/1-113
      YOU MAY NOW SEE OTHER INFORMATION ON YOUR SEARCH RESULTS; TYPE: E SHOW
      FOR EXPLANATION OF OTHER ACTIONS YOU MAY NOW DO TYPE: E COMMANDS

      1046.3
      *****USER::

U30:  SHOW INDEX DYNAMIC

         . . .

U31:  FIND DYNAMIC

C31:  ++++CONIT:
      NO RESPONSE RECEIVED FROM LAST TRANSMISSION.
      SHOULD CONIT WAIT FOR A RESPONSE ANY LONGER?  (ANSWER YES OR NO)
      *****USER::

U32:  YES

C32:  FROM HOST:
      YOUR SEARCH RESULTED IN SET 4 WHICH CONTINS 2135 DOCUMENTS.
      TO SEE REFERENCES TO THE FIRST FEW DOCUMENTS TYPE: SHOW

      1049.1
      *****USER::

U33:  COMBINES  [User hits BREAK key after 'S']

C33:  ACTION/MESSAGES ABORTED.  CONTINUE.

      1049.6
      *****USER::

U34:  COMBINE SET 2 AND SET 4

       . . .

U53:  COMBINE SET 10 AND NOT SET 3

       . . .

U56:  PICK 73
```

```
C56:  ++++CONIT:
      LOGGED OFF DIALOG.  COSTS FOR THE USE OF LAST DATA BASE:
       $64.12  0.916 HRS FILE7 5 DESCRIPTORS
        $7.33  TYMNET
       $12.10  121 PRINTS
       $83.55  ESTIMATED TOTAL COST
      TIME OF LOGOFF 8:18:24
      LOGGING ONTO SDC/ORBIT.
      FULL RESPONSE NOT YET RECEIVED FROM RETRIEVAL SYSTEM.
      SHOULD CONIT WAIT FOR A RESPONSE ANY LONGER?  (ANSWER YES OR NO)
      *****USER::

U57:  YES

C57:  NEWS FROM SDC/ORBIT:
      SAE DATA BASE NOW OPERATIONAL.  SEE NEWS RE FREE INTRODUCTORY TIME.
      *****
      YOU ARE NOW CONNECTED TO THE SDC/ORBIT RETRIEVAL SYSTEM--
      YOU ARE NOW CONNECTED TO THE SSIE (NUMBER     DATA BASE.
      FOR EXPLANATION OF FINDING DOCUMENTS TYPE: E FIND

      1117.9
      *****USER::

       . . .

U78:  SHOW SET SHOW SET 5 D1-5

C78:  ++++CONIT:
      A number should appear where you typed "show set 5 di-5"
      For an explanation of how to use the SHOW command type:  e show

      1136.4
      *****USER::

U79:  SHOW SET 5 D1-5

       . . .

U84:  S REVIEW

C84:  ++++CONIT:
      set 1:  ALL MODEL:  (43383)
      set 2:  ALL WORLD:  (46581)
      set 3:  1 AND 2  (606)
      set 4:  ALL DYNAMIC: (11555)
      set 5:  2 AND 3  (133)
      set 6:  5 AND NOT 3  (79)
      set 7:  P   (4853)
      set 8:  ALL SOCIAL:  (37749)

      1144.1
      *****USER::
```

```
U85:  COMBINE SET 8 AND SET 3

C85:  ++++CONIT:
      Your search resulted in set 9 which contains 198 documents.
      To see references to the first few documents type:  show

      1144.6
      *****USER::

U86:  SHOW

C86:  ++++CONIT:

      -1-

      AN - GY 68729 1
      TITLE:  MODELING REGIONAL GROWTH FOR EFFECTIVE AGRICULTURAL PLANNING
              AND ITS IMPLEMENTATION
      IN  -   LIANG, T ; VIETH G
      OS  -   UNIV. OF HAWAII, AGRICULTURAL EXPERIMENT STA., AGRICULTURAL
              ENGINEERING, GILMORE HALL, ROOM 102, HONOLULU, HAWAII, 96822
      PP  -   10/76 TO 9/77
      FY  -   77
      SPO -   Hawaii State Government, Honolulu, Hawaii
      CN  -   0068729; HAWOO523-S

                . . .

      -4-
      AN  -   GY 43458

U87:  [User hits BREAK key while output is coming from remote host]

C87:  Action/messages aborted; you may continue when you get the USER cue.

      1145.8
      *****USER::

       . . .

U89:  SHOW ALL SET 9 D1-198 OFF
```

```
C89:    ++++CONIT:
        Sending mail address

        YOUR OFF-LINE PRINT NUMBER IS P0084524.
        OFF-LINE PRINT COMPLETED.

        You may now see other information on your search results; type: e show
        For explanation of other actions you may now do type:  e commands

        1147.0
        *****USER::

U90:    STOP

C90:    ++++CONIT:
        End of SOC/ORBIT session.  Time statistics:
        TERMINAL SESSION FINISHED 12/20/77   8:45 A.M.  (PACIFIC TIME)
        ELAPSED TIME ON SSIE:  0.50 HRS.
        TOTAL ELAPSED TIME:  0.51 HRS.

        Logged off SOC/ORBIT.
        The phone will be hung up.
        Phone hung up.
        Phone line detached
        Thank you for using CONIT.
        12/20/77   1147.8 est Tue
        You may hang up your connection now.
```

APPENDIX C

OFFLINE INSTRUCTION MATERIAL

The materials reproduced in this Appendix were presented to the experimental users as explained in Section 3. The briefing was read to each EU. The other materials were given in the form of a printed handout to the three EU's from Group A.

List of Materials

1. Briefing
2. How to Develop a Search Strategy (two pages)
3. Attachment 1: Work Sheet for Search Terms and Databases (one page, as filled out by EU1)
4. Attachment 2: Professional Fields and Corresponding Databases (seven pages)
5. Attachment 3: Database Descriptions (one page of 14 in Attachment 3)
6. Attachment 4: Alphabetical list of Databases, by CONIT Name and Number (one page)

PRE-SESSION BRIEFING FOR EU'S

We are attempting to design a computer interface that will enable end users to do their own searching in machine-stored bibliographic databases. We want your help in evaluating an experimental interface called CONIT that allows you to communicate with three different Information Storage and Retrieval Systems located in various parts of the country.

Our purpose in this experiment is to determine whether or not CONIT has all the features that are needed to enable you to do your own searching. We also want to know about the undesirable or inconvenient features of the system as well as the good ones.

Although one or more experimenters will be on hand to help you out of difficulties, we would like to see if you can master the system by yourself and extract useful information from it. Please ask for help only if you feel hopelessly hung up.

We shall also appreciate your staying with us for a debriefing session after you finish searching. How much time do you have today?

To assist us in interpreting results of the experiment, kindly let us know about any special reactions and impressions you have as you progress. You can express them orally, or note them on a pad or on the printout paper.

Printouts of references can be obtained online or offline. We request that you limit the offline references to a total of 200. These will be furnished to you free of charge when we get them, in about a week. We also plan to do our own search on your problem. Any new documents that seem relevant will be sent to you.

Please do not be afraid that you may damage the system; it is essentially damage-proof.

We are required by law to ask your permission to use the results of this experiment in our report. We shall identify participants by their EU numbers only. May we have your permission? You are, of course, free to discontinue the experiment at any time.

[The following was read to EU's in Group A only.]

Here are instructions for developing a search strategy. When you think you have a good set of search terms and databases in which to search, please begin working at the console. Kindly leave the list of search terms you have compiled when you leave.

HOW TO DEVELOP A SEARCH STRATEGY

Before using CONIT, you should plan a search strategy. Here are some suggestions:

1. Compile a List of Search Terms

Write down a list of words that describe the topic on which you are seeking information; a work sheet is attached for your convenience. Your list should include broad, general terms [such as the professional discipline(s) or subdiscipline(s) to which your problem is related], less general terms, and narrow, specific terms. Break your search topic into concepts and list words that characterize the concepts. Try also to think of synonyms for the words you write down. This list should help you get started, but it may turn out that additional terms will show up as you proceed.

We want to emphasize strongly the importance of listing many different words and phrases that describe your problem.

2. Selection of Systems and Data Bases

Enter on your work sheet the data bases that are likely to contain documents pertaining to your search terms. An alphabetized list of professional fields and data bases that cover the fields is attached. Also appended is a detailed description of each data base. Try to match data bases to your search terms. Bear in mind that, for completeness, you may want to search more than one data base; the computer will explain how to change data bases.

3. Hints on Searching

Getting Started

Start with the search term you think will most likely yield useful documents. In general, it is preferable to use single-word terms that indicate specifically the key concepts in your topic. You can ask the computer to explain how to combine result from your single-word searches. Instructions on how to make your request are available in CONIT.

Documents are often indexed, in part, from a carefully controlled classification scheme and a controlled vocabulary. To get started, therefore, you may have to guess what some of these terms are. Sometimes your search term will yield no documents only because its form differs slightly from the allowable one. For example, a search on 'automatization' may be negative because the allowable term is 'automation'. To determine if your term is close to an

acceptable term, you may ask for a display of index terms in the data base that are alphabetically near your term. Your term will then be placed in alphabetical relationship to allowable terms, such as in the example just cited, automated, automation [automatization], automaton ... The command for accomplishing this is: show index followed by the search term you want placed in alphabetical context.

Once a few documents are found, you may discover additional good search terms by looking at all the information available for each document. The command is: show all. Details of how to use this command will be available when you get online.

No Documents Found

If no documents are found, try searching on synonymous terms, or try broader terms, or look for valid terms with the show index command.

Too Many Documents Found

If too many documents are found, most of which appear irrelevant, you may reduce the number of irrelevant documents in these ways:

a) By searching on narrower, more specific terms
b) By following a search that yields many irrelevant documents with one or more searches on narrow specific terms, and then combining the new sets with the first set. You can ask the computer to give an explanation of how to do this.
c) By searching on two-word and three-word phrases.

A final hint: Be flexible as your searching progresses. There are no hard and fast search-strategy rules. Be prepared to alter your search plan as you progress.

When you feel you have an initial search plan, begin your searching at the CONIT terminal. Step-by-step instructions will be given to you by CONIT.

---

attachments:
(1) Work sheet
(2) List of professional fields
(3) Description of data bases
(4) Alphabetical List of Data Bases

Attachment 1

WORK SHEET
FOR
SEARCH TERMS AND DATA BASES

Search Terms	CONIT Designations for Data Bases	
	Number	Name
	56a	ED
Libraries	53	PSYCH
librarians	56b	LIBINFO
public libraries	51	Socsci
public librarians	? 71	NTIS
library reference services	54c	PUBLIC
social service agencies	52	SOC
human service agencies		
public service agencies		
community service agencies	1 = public	
community information services	2 = libraries	
community information sources.	3 = 1+2	
	4 = community	
	5 = service	
	6 = 4+5	
	7 = cooperation	
	8 = 3+6	
	9 = 7+8 (10 docs)	
CONIT SET1 AND SET2		
2. LIBRARIES	10 = 2+9	15 = intersection
6 = SOCIAL	11 = public	16 = 14+15
7 2+6	12 = 2+11	17 = cooperation
10/11/77 8 = services	13 = 10+12	18 = 14+17
9 = 1+8	[illegible]	(30 docs)

Attachment 2

PROFESSIONAL FIELDS AND CORRESPONDING DATA BASES

A

Accounting
- (C) ACCOUNTANT, 62b
- (G) BUS, 61; MANAGE, 62a

Acoustics
- (G) PHYS, 12; EE, 22; ENGIN, 21
- See also: Physics

Aeronautics
- (G) ENGIN, 21; NTIS, 71; SCI, 11

Agriculture
- (C) AGRICON, 41b; AGRIRES, 41r; AGRIUS, 41a
- (S) FOOD, 41c
- See also: Business, Engineering, General, Nutrition, Science

Air Pollution
- (C) AIRPOLLUT, 35b
- (G) POLLUT, 35a; NTIS, 71; ENVIR, 34
- See also: Environmental Science

Aluminum
- (C) ALUM, 24b
- (G) METALS, 24a; ENGIN, 21

Animal Science
- See: Life Sciences, Agriculture

Anthropology
- (G) SOCSCI, 51; ANTHIST, 57b; SOC, 52; BIOSIS, 42
- See also: Social Sciences

Aquatic Sciences
- (C) OCEAN, 33a; AQUA, 35b
- (S) BIOSIS, 42

Archaeology
- (G) SOCSCI, 51
- See also: History

Art
- (C) ART, 58

Artificial Intelligence
- (G) EE, 22; NTIS, 71; ENGIN, 21; ME, 23, SOCSCI, 51 LANG, 55a

Astronomy
- (G) PHYS, 12; SCI, 11; NTIS, 71
- (S) METEOR, 32

Astrophysics
- (G) PHYS, 12; SCI, 11; NTIS, 71
- (S) METEOR, 32

Atmospheric Sciences
- (G) NTIS, 71; METEOR, 32
- (S) AIRPOLLUT, 35b
- See also: Earth and Space Sciences

Audio-visual Materials
- (S) EDAV, 56c; MEDAV, 43b

B

Banking
- (G) MANAGE, 61; FED, 54a; BUS, 3
- See also: Economics

Biochemistry
- (G) MEDLINE, 43; CHEM77, 13a; CHEM72, 13b; CHEM70, 13c; CHEMIN, 131; BIOSIS, 42; SCI, 11; NTIS, 71
- See also: Medical Sciences, Chemistry, Life Sciences

Bioengineering
- (G) MEDLINE, 43; ENGIN, 21; NTIS, 71; BIOSIS, 42; EE, 22
- See also: Life Sciences

Biology
- (C) BIOSIS, 42
- See also: Life Sciences, Dictionaries (p. 7)

Biometrics
- (G) BIOSIS, 42
- See also: Life Sciences

Biophysics
- (G) BIOSIS, 42; PHYS, 12
- See also: Life Sciences

Botany
- (G) BIOSIS, 42; AGRIUS, 41a; AGRICON, 41b; AGRIRES, 41r
- See also: Life Sciences

Business
- (C) BUS 61
- (S) ACCOUNTANT, 62b; CHEMNEWS, 66a; DRUGNEWS, 66b; MARKETAB, 63a; MARKETIN, 631; MARKETHOW, 63a; OILNEWS, 37w; STATUS, 64b; STATUSTIME, 64c; STATREG, 64d; STATFRN, 64f; STATFRNTIME, 64g; STATPLANT, 64a
- See also: Economics, Management, Special Data Bases (p. 7).

C = data bases that are central to the discipline

G = data bases that cover the discipline and others, too

S = data bases for a specific subdiscipline

Professional Fields and Corresp. Data Bases

C

Cancer
(C) CANCER, 46; CANCERRES, 46r
See also: Medical Sciences

Cardiology
See: Medical Sciences

Chemistry
(C) CHEM77, 13a; CHEM72, 13b; CHEM70, 13c; CHEMIN, 13i;
(G) SCI, 11; MEDLINE, 43; NTIS, 71
(S) CHEMNEWS; 66a; PATCHEM77, 26a; PATCHEM50,26b; PATCHEMCA 26c; PAPER, 26p; TOX, 44; TOXEFF, 44d
See also: Dictionaries (p. 7), Medical Sciences, Patents

Chemical Engineering
(G) ENGIN, 21
See also: Chemistry

Child Abuse
(C) CHILDAB, 52b
See also: Education

Climatology
(G) METEOR, 32; ENVIR, 34
See also: Earth & Space Sciences

Clinical Sciences
See: Medical Sciences

Communications
See: Engineering or Library and Information Services, Linguistics

Computational Linguistics
(G) EE, 22; SOCSCI, 51; NTIS, 71; LANG, 55a

Computer Science
(C) EE, 22
See also: Electrical Engineering, Engineering

Construction
(G) ENGIN, 21; NTIS, 71; ME, 23
See also: Business

Contracts
See: Contracts and Grants, p. 7

Counseling
(G) ED, 56a; PSYCH, 53
See also: Education, Psychology

Current Events
(C) FEDNOW, 54n; FED, 54a
(S) CRECORD, 54b; CHEMNEWS, 66a; DRUGNEWS, 66b; OILNEWS, 37w
See also: Political Science

D

Demography
(G) SOCSCI, 51; SOC, 52; STATUS, 64b; STATUSTIME, 64c

Dermatology
See: Medical Sciences

Dissertations
(C) THESES, 74

Drugs
See: Pharmacology

E

Earth and Space Sciences
(C) METEOR, 32
(G) SCI, 11; NTIS, 71
(S) GEOREF, 31; OCEAN 33a; AQUA, 33b; ENVIR, 34
See also: Climatology, Environmental Science, Geochemistry, Geology, Oceanography, Soil Science, Space Science

Economics
(C) BUS, 61; MANAGE, 62
(G) PUBLIC, 54c; SOCSCI, 51; NTIS, 71, HIST, 57a; FED, 54a; ADMIST, 57b
See also: Business

Education
(C) ED, 56a; VOTECH, 56d
(G) SOCSCI, 51; PSYCH, 53; LANG, 55
(S) EDEXCHILD, 56b; EDAV, 56c
See also: Patents

Electrical Engineering
(C) EE, 22
(G) ENGIN, 21; SCI, 11; NTIS, 71; ME, 23
See also: Engineering, Business, Patents

Electronics
See: Electrical Engineering

Energy
(C) ENERGY, 36; OIL, 37a; OILNEWS, 37w
(G) ENGIN, 21; ME, 23; ENVIR, 34
See also: Physical Sciences

Engineering
(C) ENGIN, 21
(G) SCI, 11; NTIS, 71
(S) EE, 22; ME, 23; METALS, 24a; ALUM, 24b
See also: Bioengineering, Physics, Chemistry, Business

C = data bases that are central to the discipline
G = data bases that cover the discipline and others, too
S = data bases for a specific subdiscipline

Professional Fields & Corresp. Data Bases

Entonology
(G) BIOSIS, 42
See also: Life Sciences

Environmental Science
(C) ENVIR, 34
(G) SCI, 11; NTIS, 71; MARKETAB, 63a; ENGIN, 21; ENERGY, 3b; AGRIUS, 41a; ME, 23; AGRICOM, 41b;
(S) BIOSIS, 42; GEOREF, 31; OCEAN, 33a; POLLUT, 35a; AIRPOLLUT, 35b; AQUA, 33b; METEOR, 32

Epilepsy
(C) EPIL, 45
See also: Medical Sciences

Ethnography and Ethnology
(G) SOCSCI, 51
(S) HIST, 57a; AMHIST, 57b

Ethology
(G) PSYCH, 53
See also: Life Sciences

F

Fish
(G) OCEAN, 33a; AQUA, 33b; BIOSIS, 42
See also: Agriculture, Life Sciences

Food Technology
(C) FOOD, 41c
(G) BIOSIS, 42; ENGIN, 21
See also: Nutrition

Forestry
(G) AGRIUS, 41a; AGRICOM, 41b; AGRIRES, 41r
See also: Earth and Space Sciences

Foundations, Charitable - page 7

G

General (All Fields)
(C) NTIS, 71; LIBCONF, 72e; LIBCONF, 72f; THESES, 74

Genetics
(G) BIOSIS, 42; MEDLINE, 43
See also: Life Sciences

Geochemistry
(G) GEOREF, 31; OIL, 37a
See also: Chemistry, Earth & Space Sciences, Geology

Geography
(G) SOCSCI, 51
See also: Social Sciences

Geology
(G) GEOREF, 31
See also: Earth & Space Sciences

Geophysics
(C) GEOREF, 31
See also: Earth & Space Sciences, Physics

Geosciences
See: Earth & Space Sciences, Physical Sciences

Grants
See: Contracts and Grants, p. 7

H

History
(S) HIST, 57a; AMHIST, 57b
See also: the professional field about which historical information is being sought

Horticulture
See: Agriculture

Humanities
(S) HIST, 57a; AMHIST, 57b; ART, 58; LANG, 55a; LIBINFO, 55b

Hydrology
(G) METEOR, 32; NTIS, 71
See also: Earth & Space Sciences

I

Immunology
(G) BIOSIS, 42; MEDLINE, 43
See also: Life Sciences

Industrial Engineering
(G) ENGIN, 21; ME, 23; EE, 22; NTIS, 71
See also: Business

Information Science
(C) LIBINFO, 55b
(G) SOCSCI, 51; EE, 22; NTIS, 71; ED, 56a

Insects
See: Entomology

Instrumentation Technology
(G) ENGIN, 21; EE, 22; ME, 23; NTIS, 71

International Affairs
(C) PUBLIC, 54c
See also: Political Science

C = data bases that are central to the discipline
G = data bases that cover the discipline & others, too
S = data bases for a specific subdiscipline

Professional Fields & Corresp. Data Bases

J

K

L

Law
(G) FED, 54a; FEDNOW, 54n; CRECORD, 54b; BUS, 61
See: Political Science and also the professional fields about which legal information is being sought.

Library and Information Services
(C) LIBINFO, 55b
(G) SOCSCI, 51; ED, 56a; NTIS, 71; EE, 22; AGRIUS, 41a

Life Sciences
(C) BIOSIS, 42
(G) AGRIUS, 41a; AGRICOM, 41b; AGRIKES, 41r; MEDLINE, 43; SCI, 11; NTIS, 71; SOC, 52
See also: Medical Sciences

Linguistics
(C) LANG, 55a
(G) SOCSCI, 51; ED, 56a
See also: Computational Linguistics

Literature
(G) LANG, 55a;SOCSCI, 51

M

Management
(C) MANAGE, 62; BUS, 61
See also: Business, Economics

Materials Science
(G) PHYSICS, 12; NTIS, 71
(S) ALUM, 24b; METALS, 24a; PAPER, 26p
See also: Chemistry, Engineering

Mathematics
(G) SCI, 11
See also: Accounting, Computer Science, Statistics.

Mechanical Engineering
(C) ME, 23
(S) PATENTS, 25
See also: Engineering

Mechanics
(G) PHYS, 12; ME, 23
See also: Physics

Medical Sciences
(C) MEDLINE, 43; MEDNOW, 43a
(G) BIOSIS, 42; SCI, 11; NTIS, 71
(S) CANCERLIT, 43v; MEDNOW, 43a

Medical Sciences (continued)
See also: Cancer, Epilepsy, Dictionaries (p. 7), Nutrition, Pharmacology, Psychology, Toxicology

Metallurgy
(C) METALS, 24a

Metals
(C) METALS, 24a
(G) ENGIN, 21; PHYS, 12
(S) ALUM, 24b

Meteorology
(G) METEOR 32
See also: Earth and Space Sciences

Microbiology
(G) BIOSIS, 42; MEDLINE, 43
See also: Life Sciences

Mining
(G) ENGIN, 21
See also: Metals

Molecular Biology
(G) BIOSIS, 42
See also: Life Sciences

Motor Vehicles
(G) ENGIN, 21; ME, 23; SAE, ?

N

Natural Resources
See: Environmental Science

Naval Technology
(G) ENGIN, 21; OCEAN, 33a; NTIS, 71; ME, 23

Neurology
(S) EPIL, 45
See also: Medical Sciences

Noise Pollution
(G) POLLUT, 35a; ENVIR, 34; NTIS, 71; AGRIUS, 41a
See also: Environmental Science

Nuclear Science and Engineering
(G) PHYS, 12; ENGIN, 21; EE, 22; NTIS, 71
See also: Physics

Nucleonics
(G) PHYS, 12
See also: Physics

C = data bases that are central to the discipline
G = data bases that cover the discipline & others, too
S = data bases for a specific subdiscipline

Professional Fields & Corresp. Data Bases

Nutrition
(G) MEDLINE, 43; BIOSIS, 42; AGRIUS, 41a; AGRICON, 41b; AGRIRES, 41r; [illegible], 43n
See also: Food Technology, Medical Sciences

O

Oceanography
(C) OCEAN, 33a; AQUA 33b
(G) NTIS, 71
See also: Earth and Space Sciences

Oil
See: Petroleum

Optics
(G) PHYS, 12
See also: Physics

Ordnance
(G) NTIS, 71; MARKETDOD, 63d
See also: Engineering

P

Paper
(C) PAPER, 26p
See also: Chemistry, Forestry

Patents
(C) PATDER, 25a; PATIFI, 25b
(S) PATCHEM77, 26a; PATCHEM50, 26b; PATCHEMCA, 26c
See also: Dictionaries, p. 7

Pathology
(G) BIOSIS, 42
See also: Medical Sciences

Podagogy
See: Education

Petroleum
(C) OIL 37a; OILNEWS, 37w
See also: Energy

Pharmacology
(G) MEDLINE, 43; TOX, 44; TOXEFF, 44d
(S) DRUGNEWS, 47w
See also: Medical Sciences

Philosophy
(G) SOCSCI, 51; HIST, 57a; ANHIST, 57b

Physical Sciences
(G) SCI, 11
See also: Chemistry, Earth & Space Sciences, Engineering, General, Physics

Physics
(C) PHYS, 12
(G) SCI, 11; NTIS, 71; BIOSIS, 42; GEOREF, 31
See also: Chemistry, Engineering

Physiology
(G) MEDLINE, 43; BIOSIS, 42
See also: Life Sciences

Phytopathology
(G) BIOSIS, 42; AGRIUS, 41a; AGRICON, 41b
See also: Agriculture

Plants
(G) BIOSIS, 42; AGRIUS, 41a; AGRICON, 41b; AGRIRES, 41r; FOOD, 41c
See also: Life Sciences

Political Science
(C) PUBLIC 54c
(G) SOCSCI, 51; HIST, 57a; ANHIST, 57b
See also: U.S. Government

Pollution
(C) POLLUT, 35a
(S) AIRPOLLUT, 35b
See also: Environmental Science, Noise Pollution, Water Pollution

Population
(G) STATUS, 64b; STATUSTIME, 64c; STATREG, 64d
See also: Demography

Power Engineering
(C) ENERGY, 36; OILNEWS, 37w
(G) EE, 22
See also: Energy

Printing and Publishing
(G) LANG, 55a; LIBINFO, 55b
See also: Engineering

Propulsion Systems
See: Aeronautics, Naval Technology, Engineering

Psychiatry
(G) MEDLINE, 43; PSYCH, 53; SCI, 11; SOCSCI, 51
See also: Medical Sciences

Psychology
(C) PSYCH, 53
(G) MEDLINE, 43; SOCSCI, 51; SCI, 11; BIOSIS, 42; NTIS, 71, LANG, 55a

Public Affairs
(C) PUBLIC, 54c
See also: Political Science

C = data bases that are central to the discipline
G = data bases that cover the discipline & others, too
S = data bases for a specific subdiscipline

Professional Fields & Corresp. Data Bases

Public Health
(G) MEDLINE, 43; SOCSCI, 51; BIOSIS, 42
See also: Medical Sciences

Q

R

Radiobiology
(G) MEDLINE, 43; BIOSIS, 42
See also: Life Sciences
Refining
See: Petroleum
Research and Development
(C) NTIS, 71; SSIE, 73
(G) SOCSCI, 51
(S) AGRIRES, 41r; CANCERRES, 46r
Robotics
(G) ENGIN, 21; EE, 22; ME, 23
See also: Artificial Intelligence

S

Science
(C) SCI, 11
See also: Life Sciences, Physical Sciences, Social Sciences, General
Social Sciences
(C) SOCSCI, 51
See also: Economics, Education, Library and Information Services, Linguistics, Political Science, Psychology, Sociology
Sociology
(C) SOC, 52
(G) PUBLIC, 54c; SOCSCI, 51; HIST, 57a, AMHIST, 57b
Soil Science
(G) AGRIS, 41a; AGRICOM, 41b; AGRIRES, 41r
See also: Earth and Space Sciences
Space Sciences
(G) BIOSIS, 42; METEOR, 32; PHYS, 12
See also: Earth and Space Sciences
Space Technology
(G) ENGIN, 21; NTIS, 71, SCI, 11
See also: Earth and Space Sciences, Space Sciences
Statistics
(G) SCI, 11; BUS, 61
(S) STATUS, 64b; STATUSTIME, 64c; STATREG, 64d; STATFRN, 64f; STATFRNTIME, 64g; STATPLANT, 64a
Surgery
See: Medical Sciences
Symbiosis
(G) BIOSIS, 42
See also: Life Sciences

T

Technology
See: Engineering
Textiles
(G) ENGIN, 21; ME, 23
See also: Chemistry
Toxicology
(C) TOX, 44, TOXFF, 44d
(G) AIRPOLLUT, 35b
See also: Medical Sciences
Trade and Commerce
See: Business
Transportation Systems
(G) ENGIN, 21; NTIS, 71; ME, 23
See also: Motor Vehicles, Aeronautics, Naval Technology

U

Urban Planning
(G) NTIS, 71; SOCSCI, 51; AMHIST, 57b; AGRIUS, 41a; AGRICOM, 41b; ENVIR, 34
See also: Political Science
U.S. Congress
(C) CRECORD, 54b
See also: U.S. Government
U.S. Government
(C) FED, 54a; FEDNOW, 54n
(G) AMHIST, 57b; STATUS, 64b; NTIS, 71; PUBLIC, 54c; MANAGE, 62
(S) CRECORD, 54b
See also: Political Science
U.S. History
(C) AMHIST, 57b
U.S. Industrial Firms
(C) STATPLANT, 64a
See also: Business
U.S. Statistics
(C) STATUS, 64b; STATUSTIME, 64c; STATREG, 64d

C = data bases that are central to the discipline
G = data bases that cover the discipline & others, too
S = data bases for a specific subdiscipline

Professional Fields & Corresp. Data Bases

V

Veterinary Sciences
(G) AGRIUS, 41a; AGRICOM, 41b; AGRIRES, 41r; BIOSIS, 42
See also: Agriculture, Life Sciences

Virology
(G) MEDLINE, 43
See also: Life Sciences

W

Water Pollution
(G) POLLUT, 35a; OCEAN, 33a; ENVIR, 34; NTIS, 71, AGRIUS, 41a; AQUA, 33b
See also: Environmental Science

Wildlife
(G) BIOSIS, 42; ENVIR, 34
See also: Animals, Fish, Life Sciences

X

Y

Z

Zoology
See: Animal Science

Special Data Bases

Contracts and Grants
(C) FEDGRANTS, 65c; GRANTS, 65a; MARKETDOD, 63d
See also: Business

Dictionaries:
Biological Terms - BIOTERM, 42t
Chemical Names and Structures - CHEMTERM, 13t
Chemical Terms - MEDCHEMTERM, 43s
Medical Subject Headings - MEDTERM, 43t
Medical Terms - MEDCHEMTERM, 43s
Patent Classification Codes - PATTERM, 25t
Toxicological Terms - TOXTERM, 44t

Foundations, Charitable
(C) FOUNDIR, 65b

C = data bases that are central to the discipline
G = data bases that cover the discipline & others too
S = data bases for a specific subdiscipline

JFR & RSH
10/28/77

Attachment 3

DATA BASE DESCRIPTIONS

Given below are descriptions of the data bases accessible through CONIT, the experimental interface to retrieval systems. The data bases are listed in order of the number by which they are referred to in CONIT.

The format of the listings is shown in the following example:

CONIT No.	CONIT Name	Code letter of system having the data base;[1] data base name (and no.) as given in the system.	Years of coverage	; Size[2] of data base as of Sept. '77	+ Size[2] and frequency of update	; Usage cost/hr of connect time	+ Cost[3] per record of off-line printout
11.	SCI	[D:SCISEARCH(34)]	1974-*;	1,700,000	+ 42,000/mo;	$70/hr	+ $0.10/rec.

(SHORT DESCRIPTION OF DATA BASE)

---

1 Code Letters: D = DIALOG; N = NLM/MEDLINE; O = ORBIT; S = SUNY/MEDLINE

2 Approximate number of documents; a data base that is no longer being updated is tagged "FIXED" after its size listing.

3. Cost is for getting all information in one document's computer record printed offline and mailed to user; partial information may cost somewhat less.

* No final date means coverage to present date.

NOTE: If the information is different for different systems, then it is given in the order in which the systems are listed.

---

10. PHYSICAL SCIENCES (PHYSICS, CHEMISTRY)

11. SCI [D:SCISEARCH(34)] 1974-; 1,700,000 + 42,000/mo; $70/hr + $.10/rec.

Covers all fields of physical, biological, and medical literature. Corresponds to Science Citation Index published by Institute for Scientific Information. Articles citing a given author or paper can be retrieved.

12. PHYS [D:INSPEC-PHYSICS(12)] 1969-; 500,000 + 7,000/mo; $45/hr + $.10/rec.

Covers the field of physics. Corresponds to Science Abstracts - A published by the IEE (Institute of Electrical Engineers, London).

13a. CHEM77 [D:CA CONDENSATES(4)] 1977-; 240,000 + 12,000/2wk; $45/hr + $.16/rec.

Covers the fields of chemistry and chemical engineering. Corresponds to Chemical Abstracts Indexes (no abstracts) published by Chemical Abstracts Services. Also indexed by CASIA terms (see data base 13i).

13b. CHEM72 [D:CA CONDENSATES(3)] 1972-76; 1,772,194 (FIXED); $35/hr + $.08/rec.

[O:CHEMCON] 1972-; 1,900,000+12000/2 wk; $60/hr + $.12/rec.

Same coverage as CHEM77 without CASIA terms.

Attachment 4

ALPHABETICAL LIST OF DATA BASES, BY CONIT NAME & NUMBER

Name	Number
ACCOUNTANT	62b
AGRIUS	41a
AGRICON	41b
AGRIRES	41r
AIRPOLLUT	35b
ALUM	24b
AMHIST	57b
AQUA	33b
ART	58
BIOSIS	42
BIOTERM	42t
BUS	61
CANCER	46
CANCERRES	47r
CHEM77	13a
CHEM72	13b
CHEM70	13c
CHEMIN	13i
CHEMNEWS	66a
CHEMTERM	13t
CHILDAB	52b
CRECORD	54b
DRUGNEWS	66b
ED	56a
EDAV	56c
EDEXCHILD	56b
EE	22
ENERGY	36
ENGIN	21
ENVIR	34
EPIL	45
FDNDIR	65b
FDNGRANTS	65c
FED	54a
FEDNOW	54n
FOOD	41c
GEOREF	31
GRANTS	65a
HIST	57a
LANG	55a
LIBCONF	72e
LIBCONF	72f
LIBINFO	55b
MANAGE	62a
MARKETAB	63a
MARKETDOD	63d
MARKETIN	63i
MARKETNOW	63r
ME	23
MEDAV	43v
MEDCHEMTERM	43s
MEDLINE	43
MEDMON	43m
MEDNOW	43n
MEDTERM	43t
METALS	24a
METEOR	32
NTIS	71
OCEAN	33
OIL	37a
OILNEWS	37w
PAPER	26p
PATCHEM50	26b
PATCHEM77	26a
PATCHEMCA	26c
PATDER	25a
PATIFI	25b
PATTERM	25t
PHYS	12
POLLUT	35a
PSYCH	53
PUBLIC	54c
SCI	11
SOC	52
SOCSCI	51
SSIE	73
STATFRN	64f
STATFRNTIME	64g
STATPLANT	64p
STATREG	64d
STATUS	64s
STATUSTIME	64c
THESES	74
TOX	44
TOXEFF	44d
TOXTERM	44t
VOTECH	56d

APPENDIX D

DETAILS OF EXPERIMENTAL SESSIONS

In this appendix, we summarize some details of the experimental sessions which are pertinent to the analysis given in the body of the report. In this summary we emphasize the problems encountered by the EU's in handling system mechanics. Search strategy considerations are discussed in Section 3.

EU1

1. EU1 came to the session with a written list of 11 terms from the ERIC thesaurus which she had found in browsing through a printed copy of the thesaurus.

2. EU1 was confused by the message CONIT gave concerning delay in host system response to several search commands. Rather than have CONIT wait for the response, she went on to give other commands, and hence lost the results of those searches. After seeing this happen twice, the supervisor suggested to the EU that she wait in those situations; the EU followed this advice subsequently in the session. After the session, that CONIT message was revised and caused no trouble for the remaining five EU's.

3. EU1 asked how to request offline output. The supervisor asked her to try to get the explanation through CONIT. The EU did find the correct explanation herself, but required four EXPLAIN commands to do it. This difficulty related, in part, to a general difficulty experienced by several of the EU's: the problem of finding previous explanations in a mass of typescript.

4. EU1 had some trouble keeping track of the set numbers for the different retrieved sets. At one point, she used a wrong set number in a COMBINE

command and this hindered her search strategy formulation somewhat.

5. EU1 looked at only standard (citation) output for the document while she was online. This was one factor in misleading her into believing she had retrieved documents of higher relevance to her topic than was actually the case in some instances.

6. EU1 stopped the online session after a relatively short period of time (50 minutes) because: (1) she was very happy to have retrieved a number of good document references as a start toward a research paper; and (2) she "didn't want to use too much computer time" (a constraint we had not intended; see briefing in Appendix C).

EU 2

1. EU2 made many (eight) spelling errors caused by poor typing. As with other EU's who made relatively few errors of this type, he was able to recover from these errors fairly quickly.

2. EU2 made a number of syntactical errors. Because he was not able to understand fully several features of the system, he could not always give a clear explanation , in the debriefing, of what he was trying to do; however, an attempt at such explanations is given below:

a) He gave "FIND" instead of "E FIND" when seeking an explanation of the FIND command.

b) He tried to use COMBINE to "save sets".

c) He used "SHOW WINGS" to try to get output from the "WINGS" search. (He did correct this in his next command.)

d) He tried to combine more than two sets at a time.

e) He typed "SHOW ALL D3 d" in order to get "type 3 information [??? Maybe Set 3]. CONIT took this to mean SHOW ALL for just

document 3 in the current set (which was set 3); perhaps, thereby, fortuitously doing what EU2 wanted.

f) He had trouble figuring out how to get offline output. Because the hour was getting late, the supervisor showed the EU how this was done.

3. EU2 preferred the unabbreviated forms of commands as easier to understand when reviewing what he had done.

4. Despite all the problems, EU2 did do all the searching and online output without human assistance and was highly pleased and satisfied with the few relevant documents thus found.

EU3

1. The retrieval system automatically logged out due to excessive time without interaction while EU3 was getting the search explanations. CONIT's message was: "The phone connection and retrieval system have been disconnected." The following dialogue ensued between the EU and the supervisor:

EU: "Did the system crash?"

SUP: "Yes, what do you think you should do to get it back?"

EU: "Pick it again".

SUP "Good!"

EU3 then picked the database again and was reconnected without additional problems.

2. The "TIME OVERFLOW" message was received by CONIT from ORBIT while waiting for the response to a search. Due to a bug in the CONIT rules

(since fixed), this caused CONIT to hang up, rather than send ORBIT the request to continue searching. The supervisor forced CONIT to do what it was supposed to have done, and then returned control to the EU.

3. This EU (3) made the error several times of omitting the space after the command name. He easily understood the problem and was able to continue quickly with the correct format.

4. The supervisor suggested termination of the online session after 110 minutes for reasons of time, and because EU3 was having some difficulty understanding the indexing in this database (see discussion in text). The supervisor permitted the EU to exceed the nominal limit of 200 offline references in order to allow the session to be completed without losing the current results.

5. The EU was very pleased with his use of CONIT: he estimated he had saved about two weeks of researching in the library.

EU4

1. EU4, as did other EU's, complained about the difficulty of looking back in the (lengthy) typescript to refer to previously printed explanations.

2. The first attempt to pick the PHYS (INSPEC) database foundered on the EU's attempt to connect to DIALOG over TYMNET. EU4 was somewhat confused by the CONIT message suggesting he pick another system, since he had picked a database (an oversight in bringing the new CONIT's messages fully into the virtual-system mode). Nevertheless EU4 was able to overcome the problem by simply reissuing his PICK request.

3. EU4 was, as were other EU's, left unsure of what to expect when CONIT said it would show "the first few" documents. (CONIT did this because its simplified translation -- at that time -- of the SHOW command to the "PRINT command resulted in a variable number of document references.)

4. EU4 noted a mistake in the CONIT explanation of the COMBINE command (the word "and" had been used instead of "or" in one example).

5. Due to a bug in its rules, CONIT asked the user if he wanted to wait after handling a user BREAK request. EU4 recovered from this confusion after a prompt from the supervisor that implicitly suggested he answer "no" to the faulty CONIT message.

6. EU4 was somewhat confused about how many times he should wait for response from a search, but this did not prevent his continuing and getting the results.

7. An unexpected and unexplained message from DIALOG,"Msg from 9050:HISPEED LINE TEST",was passed on to the EU along with the results from a search. The message did not seem to bother EU4.

8. EU4 issued the command SHOW INDEX 1, apparently expecting to see part of a thesaurus classification. He recovered quickly ("Oh, this is an index") and immediately issued the semantically correct command, "SHOW INDEX ELECTRON.

9. Due to a logical error in a CONIT rule, several of the EU's searches on the 'reference tags' (e.g., E6) for terms found from SHOW INDEX requests were done incorrectly. The supervisor, on observing this, recommended that searching only on the full spelling of terms be done. EU4 followed these directions, but his searching was hindered.

10. EU4 misread the instructions to type "SHOW MORE" to continue a SHOW INDEX output; he typed just "MORE" first, and then "SHOW MORE", but CONIT was not prepared to handle that erroneous sequence. EU4 was able to recover by redoing the original SHOW INDEX command, but this did slow down his searching.

11. EU4 was confused by the headings "Type" and "RT" in the SHOW INDEX response. (These headings were superfluous in the given context; they apply only if thesaurus information is available.)

12. EU4 was confused by the free-vocabulary component of the INSPEC database indexing. He wondered why there were not more documents posted under these terms (as there would be if they were well indexed controlled terms).

13. EU4 asked for "E FIND BETTER" (explanation for improving precision) when he probably (should have) wanted "E FIND MORE" (to improve recall).

14. The truncated search on "electron" resulted in an overflow response for which CONIT was not prepared to help the user. This failure stifled a potentially effective line of searching.

15. The communications with DIALOG were then disrupted and EU4 decided to try a different database: SCI.

16. EU4 tried to do an author search but was foiled by the variant author format in this database (no comma after last name). In fact, EU4 wanted to do a citation search, which was not yet possible in the virtual mode. He also tried unsuccessfully to use SHOW INDEX to see authors -- this does not work for DIALOG, for which one must search the author index.

17. MULTICS crashed, forcing a premature termination of the session; EU4 was disappointed with the limited results.

EU5

1. EU5 misread database "55a" as "55r" and used the wrong spelling in an EXPLAIN DATA command.

2. EU5, still not appreciating the mistake, asked to EXPLAIN DATA with the name of the database (LANG). Unfortunately, the connection to DIALOG required for this EXPLAIN did not work. This, perhaps, led EU5 to try to PICK the system DIALOG, rather than a database. He then asked for the explanation of the LANG database again, which he finally got.

3. EU5 used wrong syntax to get explanation for the SCI database (E SCI).--- although he had used the proper syntax twice before. This mistake may be linked to his next mistake in believing that he was connected to SCI when he was still connected to the default database, ERIC. This, in turn, led to his searching in ERIC for 10 minutes before realizing the mistake.

4. EU5 issued erroneous command E FIND E6, but then, intuitively figured out that it should be FIND E6.

5. Other syntactic errors were:

a) FIND SMITH (forgot AUTHOR parameter)

b) COMBINE 1 AND 2 (forgot SET parameter)

c) FIND INDEX E8 (meant FIND E8)

6. EU5 thought SHOW ALL would show all information on all documents.

7. EU5 was happy to have retrieved the documents he did, but would have liked more complete recall.

EU6

1. EU6, as did several other EU's, would have liked a printed reference manual to refer to.

2. EU6, as was true of several other EU's, did not always wait for the USER cue. These mishaps caused minor delays, but generally were easily understood and overcome.

3. EU6 searched on several problems; this prevented more in-depth searching of any one problem.

4. Late in the session, EU6 was unsure of the proper use of COMBINE, even though he had used it properly earlier. The incomplete form, COMBINE SET, led to the suggestion to E COMBINE, which in turn reminded EU6 of the correct usage.

5. EU6 was happy to have retrieved the documents he did, but would have preferred more complete recall.